KEYNOTE (GREEN)

Enhancing preclinical drug discovery with artificial intelligence

R.S.K. Vijayan [a], Jan Kihlberg [b], Jason B. Cross [a,*], Vasanthanathan Poongavanam [b,*]

[a] Institute for Applied Cancer Science, MD Anderson Cancer Center, Houston, TX, USA
[b] Department of Chemistry-BMC, Uppsala University, Uppsala, Sweden

Artificial intelligence (AI) is becoming an integral part of drug discovery. It has the potential to deliver across the drug discovery and development value chain, starting from target identification and reaching through clinical development. In this review, we provide an overview of current AI technologies and a glimpse of how AI is reimagining preclinical drug discovery by highlighting examples where AI has made a real impact. Considering the excitement and hyperbole surrounding AI in drug discovery, we aim to present a realistic view by discussing both opportunities and challenges in adopting AI in drug discovery.

Keywords: Artificial intelligence; Machine learning; Deep learning; Drug discovery

Introduction

Drug discovery is a long, complex, and high-risk process. It typically takes a staggering 10–15 years and costs up to US $2.8 billion on average, to develop a new drug, while an astonishing proportion (80–90%) of them fail in the clinic,[1] with Phase II proof-of-concept (PoC) trials accounting for the most significant number of clinical failures. Although the number of new molecular entities (NMEs) approved by regulatory agencies, such as the US Food and Drug Administration (FDA), has increased over the past decade (2010–2019) compared with the prior decade, the cost of bringing a new drug to market has risen precipitously.[1–3] The key drivers contributing to the increased cost of pharmaceutical innovation include investment lost from late-stage clinical attrition, an increasingly stringent regulatory system that sets a high bar for approval, and higher clinical trial costs, especially for

Vijayan Ramaswamy (R.S.K. Vijayan) is a research scientist with the Structural Chemistry group at the Institute for Applied Cancer Science, University of Texas MD Anderson Cancer, TX, USA. Before starting at the MD Anderson Cancer in 2016, he was a postdoctoral fellow at Rutgers University, NJ, USA, and at Temple University, PA, USA. He received his Ph.D. as a CSIR senior research fellow from the Indian Institute of Chemical Biology, Kolkata, India. He is a named co-inventor on 7 issued US patents, including an ATR kinase inhibitor that has advanced to clinical trials. He has published more than 20 scientific articles and authored one book chapter. His research focuses on applying computational chemistry methods to drive small-molecule drug discovery programs,-particularly in oncology and neurodegenerative diseases.

Jan Kihlberg holds a chair in Organic Chemistry at Uppsala University (Sweden) since 2013. Before moving to Uppsala, he spent 10 years at AstraZeneca R&D in Gothenburg, seven of which as Director of Medicinal Chemistry. He was appointed professor in bioorganic chemistry at Umeå University in 1996, after having established his independent research group at Lund University in 1991. Currently, his main research interest is to understand what properties convey cell permeability, oral bioavailability, and target binding to drugs in the beyond rule of 5 space, in particular macrocycles and PROTACs. He has published 168 scientific articles, and 16 reviews and book chapters.

* *Corresponding authors. Cross, J.B. (jbcross@mdanderson.org), Poongavanam, V. (vasanthanathan.poongavanam@kemi.uu.se).*

Jason B. Cross leads the Structural Chemistry group at the Institute for Applied Cancer Science, University of Texas MD Anderson Cancer Center, TX, USA. Before joining MD Anderson in 2015, he provided computational chemistry support for drug discovery projects at Affinium Pharmaceuticals and Wyeth Pharmaceuticals, and led the molecular modeling group at Cubist Pharmaceuticals. He received his Ph.D. from Wayne State University, MI, USA, before completing a postdoctoral fellowship at Pfizer Inc. in Ann Arbor, MI, USA. He is a named co-inventor on more than 20 issued US patents, including three compounds that advanced to clinical trials. He has published more than 20 scientific articles and three book chapters. His research focuses on the application of computational methods to support small-molecule drug discovery.

Vasanthanathan Poongavanam is a researcher in the Department of Chemistry-BMC, Uppsala University, Sweden. Before starting at Uppsala University in a postdoctoral position with Jan Kihlberg in 2016, he was a postdoctoral fellow at the University of Vienna, Austria, and at the University of Southern Denmark. He obtained his Ph.D. in medicinal chemistry as a Drug Research Academy Fellow at the University of Copenhagen, Denmark, on computational modeling of Cytochrome P450. He has published more than 45 scientific articles, including reviews and book chapters. His research interests focus on understanding the molecular properties that govern the pharmacokinetic profile of molecules beyond the Ro5 space, including macrocycles and PROTACs.

pivotal trials. Given these realities, pharmaceutical and biotech companies are incentivized to innovate and adopt new technologies to improve productivity, cut costs, and ensure sustainability.

AI is shaping the evolution of entire industries, including health care (IBM Watson Health and Google's DeepMind Health). Unsurprisingly, the biopharmaceutical industry also recognizes the potential value of AI and has shown keen interest in adopting AI-driven discovery platforms in the hope of streamlining R&D efforts, reducing discovery timelines and cost, and improving efficiency.[4–5] Major pharmaceutical companies have made significant investments in AI technology, including equity investments, acquisitions of, or partnerships with, AI-focused companies, building internal capabilities, or a combination of approaches. Partnerships appear to be focused on fast tracking the development of novel therapies, drugging 'undruggable' targets, broadening target portfolio through identification of novel targets, and improving the odds of clinical success.[6] Big-tech companies, such as IBM, Microsoft, Amazon, and Google, which have competency and expertise in AI, are also making a foray into the drug discovery space.[7] Public–private initiatives, such as the ATOM consortium (https://atomscience.org), have also been established with a mission of transforming drug discovery using data-driven modeling.

The AI technologies used today in drug discovery have evolved from earlier machine learning (ML) and cheminformatics concepts. For example, the application of ML to the development of quantitative structure–activity relationship (QSAR) models and expert systems for toxicity prediction has a longstanding history.[8–9] The widespread adoption of these techniques witnessed in recent times has been fueled by the advent of big data, advanced analytics, GPU-acceleration, cloud computing, algorithm development, and the democratization of AI toolkits.[4]

The use of AI technologies is driving new opportunities across the drug discovery and development continuum, starting from target identification through to preclinical development (Fig. 1a). Evidence suggests that lack of clinical efficacy has been the foremost cause of attrition in clinical Phase II studies,[10] highlighting that target selection remains one of the most crucial decision points in drug discovery. Consequently, there is a desire to improve the target selection process by applying AI techniques. AI-driven target discovery platforms can extract and synthesize target-relevant information from a large volume of complex, disparate multi-omics data, providing a better understanding of target biology, uncovering disease–target associations, thus identifying targets with a strong link to a disease. TargetDB is one such example that integrates publicly available data on a given target and uses an ML-based classification system to categorize target tractability.[11] The approach and scoring system used within TargetDB provides useful criteria for ligandability assessment and prioritization of drug targets for development.

Once a target of interest has been identified and validated, the next stage in drug discovery is to identify high-quality chemical start points (hits) that bind to, and modulate the target. Although there are a range of hit-finding methods available, virtual screening (VS) is a cost-effective, and resource-sparing approach used to prioritize a subset of compounds for evaluation in a primary assay. The use of AI-driven approaches to improve the performance of VS is increasing.[12] AI-powered VS campaigns have identified novel hits against seemingly difficult to drug targets,[13–14] thereby turning them into tractable drug targets.

To ensure that quality hits worthy of further consideration are progressed, computational methods have been used to identify, prioritize, and select hit compounds, a process referred to as hit triage.[15] ML models are now being used to automate and improve the efficiency of this process.[16]

Fast, accurate, and reliable prediction of binding free energies to enable VS and structure-based design remains a significant challenge, including rank-ordering of compounds from a VS. In recent years, ML-based scoring functions

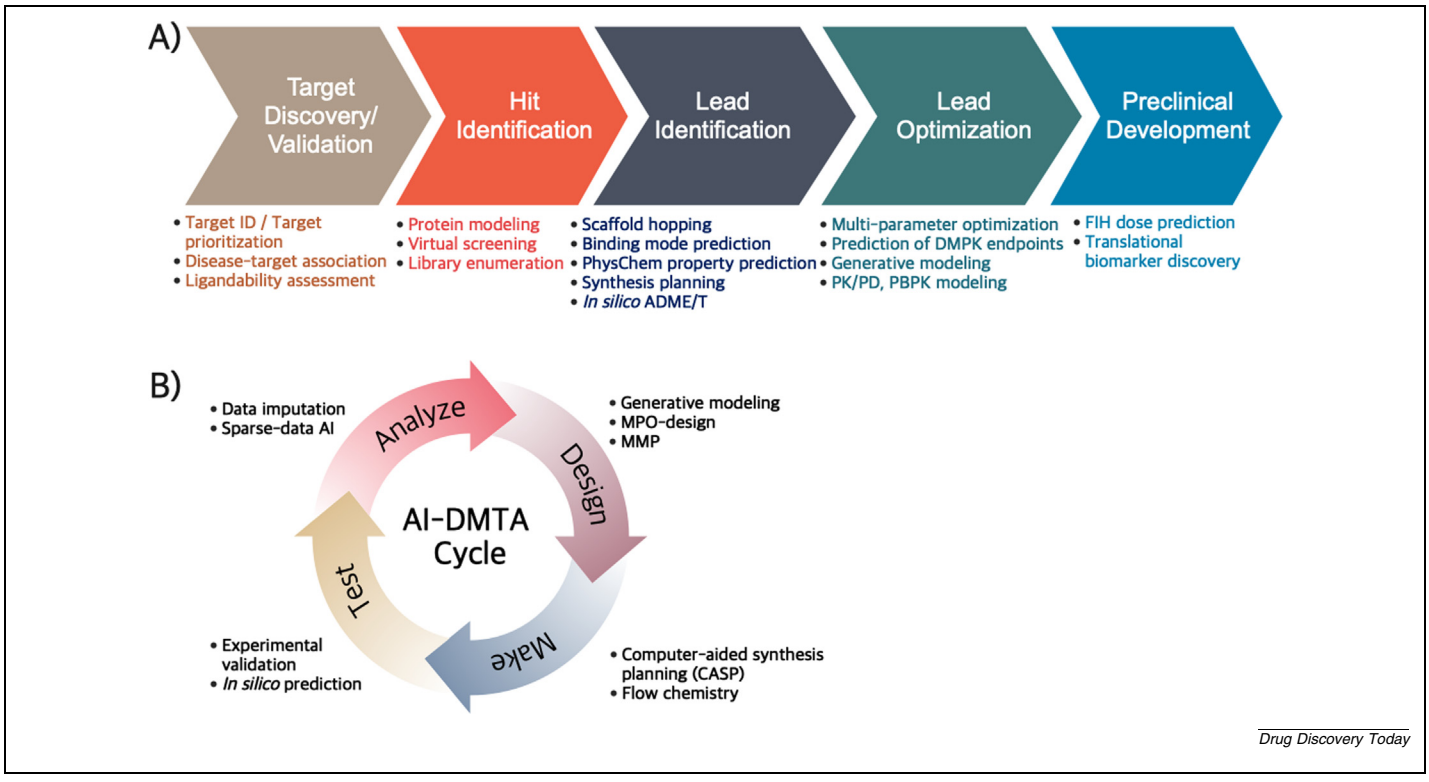


FIGURE 1

Opportunities for the application of AI techniques in drug discovery. **(a)** Schematic representation of the preclinical drug discovery process highlighting opportunities for AI to be applied across the drug discovery continuum. **(b)** Illustration of an AI-driven design–make–test–analyze cycle, highlighting opportunities to leverage AI to increase the effectiveness of the DMTA cycle.

trained on databases of protein–ligand complexes have shown great promise in improving hit rates during VS.[17] Unlike traditional scoring functions, ML-based scoring functions can implicitly account for molecular interactions that are difficult to model, and are not constrained to any predefined functional form. With the advent of 'make-on-demand' libraries and screening collections breaking the billion compound barrier, conventional docking methods have become impractical. Active learning methods integrated with molecular docking offer an elegant solution for efficient exploration of chemical space through iterative screening.[18–19]

The lead optimization (LO) phase is the most expensive and time-consuming phase in preclinical drug discovery.[20] It is inherently a multiparameter optimization (MPO)[21] problem, with the goal of identifying compounds with an optimal balance of drug-like properties while maintaining sufficient potency. Hitting this 'sweet spot' is a challenge, because it involves simultaneous optimization of multiple and often competing objectives, such as safety, specificity, efficacy, and pharmacokinetics (PK) properties, while maintaining potency.[22]

LO involves iterative rounds of the design–make–test–analyze (DMTA) cycle (Fig. 1b), and reducing the number of iterations is crucial for accelerating the LO process. Generative chemistry that relies on AI-guided generative modeling for compound design has demonstrated success in reducing the number of iterations and designing compounds that meet the defined LO criteria.[23] Generative modeling platforms also integrate various predictive models for calculating various absorption, distribution, metabolism, excretion, and toxicity (ADMET) endpoints to guide the design and selection of compounds with favorable properties that satisfy the defined LO criteria. In this way, generative chemistry can automate and shorten the 'design' phase of the DMTA cycle and offset individual cognitive biases during molecule ideation.

AI is also making headway in computer-aided synthesis planning (CASP), which is valuable in both hit identification and improving DMTA cycle efficiency.[24] AI-assisted synthesis planning helps chemists to objectively choose the most efficient and cost-effective synthetic route for a target molecule, thus accelerating the 'make' phase of the DMTA cycle. Automated continuous flow chemical synthesis is another emerging technology poised to revolutionize organic synthesis.[25] This technology opens new avenues by integrating smart automation and intelligent synthesis, thereby enabling fully autonomous synthesis.

Closing the loop of the DMTA cycle is the 'analysis' phase. To improve DMTA cycle efficiency, the data must be turned into knowledge to make better design suggestions for the next iteration. Given the sparse and non-uniform nature of the data encountered in drug discovery, the incorporation of sparse data AI methods, such as few-shot learning, for data analysis allows extracting valuable insights to inform the next round of the design cycle. Another practical application of AI is using deep imputation methods to handle the noisy, sparse, missing, and truncated data often generated in drug discovery.[26–28] Deep imputation methods combine DL and statistical imputation

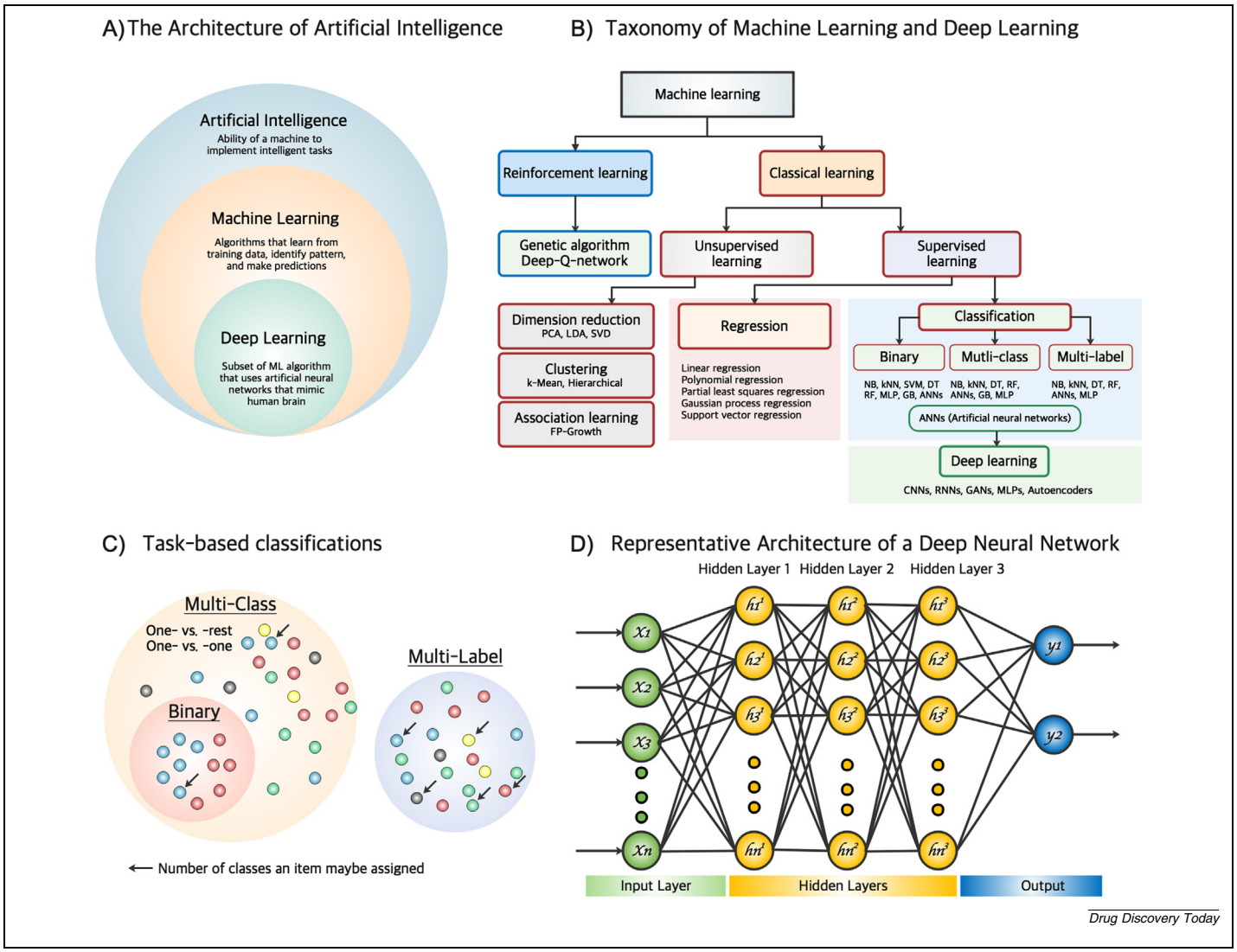


FIGURE 2

Relationships between terminology and techniques described in this manuscript. **(a)** Hierarchy of artificial intelligence. **(b)** Classification of machine learning and deep learning methods **(c)** illustration of task-based classification. **(d)** A simple schematic diagram of a deep neural network. Expanded abbreviations are provided in the main text.

methods to learn correlations between experimental endpoints and gain valuable information, even from minimal experimental data, to more accurately fill in missing experimental values.[28] Such techniques can help establish assay–assay correlations and build multitarget QSAR models, which can be used for *in silico* off-target profiling against protein target families, such as kinases.

Translating preclinical discoveries into clinical practice in the form of new therapeutics is one of the biggest challenges in clinical development and, too often, clinical candidates are lost during translation. To bridge this 'translational gap', translational strategies are increasingly being integrated as early as LO to improve Phase II and Phase III clinical success rates, which is most evident in oncology drug discovery programs.[29] To that end, translational biomarkers that demonstrate target modulation, target engagement, and confirm proof of mechanism (PoM) are used for de-risking clinical development. The ability of AI techniques to learn hidden and meaningful patterns by integrating large amounts of heterogeneous and high-dimensional omics data sets provides valuable insights for translational biomarker discovery.[30] As innovations in AI technologies continue, the use of AI in drug discovery will also continue to grow.

AI toolbox for drug discovery

AI draws inspiration from diverse disciplines and brings together many technologies, such as ML, deep learning (DL), and data analytics. This has also led to a continually growing vocabulary used to describe AI. Although these terms are often muddled and used interchangeably, they have distinct meanings and relationships to one another, including properties such as data requirements, complexity, transparency, and capabilities. In general, AI is an umbrella term and considered a superset of ML, which itself is a superset of DL (Fig. 2a). Table 1 summarizes AI algorithms and their application in preclinical drug discovery.

TABLE 1

ML and DL algorithms and their application in preclinical drug discovery.

AI category	Algorithms	Task	Drug discovery applications
ML/supervised	Decision tree	Classification[a,b,c]	Binding affinity prediction; binding pose prediction; VS; PhysChem property prediction; ADMET
	SVM	Classification/regression[a]	
	RF	Classification/regression[a,b,c]	
	kNN	Classification[a,b,c]	
	Naïve Bayesian	Classification[a,b,c]	
	Multilayer perception	Classification[a,b,c]	
	GB	Classification/regression[a,b]	
	ANN	Classification/clustering[a,b,c]	
	Linear regression (e.g., MLR)	Regression	
	Polynomial regression	Regression	
	Partial least squares	Regression	
	Support vector regression	Regression	
ML/unsupervised	PCA	Clustering/dimensionality reduction	VS; data visualization; dimensionality reduction; molecular pattern recognition
	k-mean clustering	Clustering	
	SOM	Clustering	
	Hierarchical clustering	Clustering	
	t-SNE	Clustering	
	Singular value decomposition	Dimensionality reduction	
	Linear discriminant analysis	Dimensionality reduction	
	Multidimensional scaling	Dimensionality reduction	
	Partial least squares regression	Dimensionality reduction	
	Principal component regression	Dimensionality reduction	
	Gaussian mixture models	Probability distribution	
	Association rule learning (e.g., FP-Growth)	Pattern recognition	
ML/RL	Q-Learning	Generative modeling	*De novo* molecular design; VS; docking; MPO
	GENTRL		
	ReLeaSE		
DL	CNNs	Representation-learning techniques	Protein structure prediction; binding affinity; binding pose prediction; *de novo* molecular design; focused library generation; PhysChem property prediction; ADMET; MPO; synthesis planning
	RNNs		
	Deep belief networks (DBN)		
	Stacked autoencoders		
	LSTMs		
	Deep Boltzmann machine (DBM)		

[a] Binary classifier.
[b] Multiclass classifier.
[c] Multilabel classifier.

ML algorithms are designed to build models that learn from problem-specific training data by identifying complex patterns and predicting outcomes on unseen data without being explicitly programmed. For these reasons, they are widely used in preclinical drug discovery[7] and have been successfully used to predict bioactivity, ADMET related endpoints, and physicochemical properties with increased levels of accuracy.[31] ML algorithms are broadly categorized as supervised, unsupervised, and reinforcement learning (RL) (Fig. 2b). The choice of the ML algorithm depends on many factors, including the data set and the type of problem.

Supervised learning methods use labeled data to train models and, once trained, the models can be used to predict outcomes on unseen data. These algorithms can handle both categorical and continuous data and are commonly used for classification and regression-based modeling approaches.[32] Classification methods can be further categorized into binary, multiclass, and multilabel depending upon the specific classification tasks (Fig. 2c). Supervised learning methods have been shown to outperform inductive learning methods in multiple instances and have been used for similarity searching, predicting bioactivity, and other properties of interest. Kernel methods, such as support vector machines (SVMs), can map high-dimensional vector spaces, allowing for molecular featurization beyond classical molecular descriptors and fingerprints (e.g., MACCS keys and extended connectivity fingerprints).[33] Supervised ML algorithms can also handle high-dimensional data to overcome the 'curse of dimensionality', and the collinearity problems often encountered in QSAR modeling,[33–35] in addition to providing options for hyperparameter optimization when searching for the best model.[36] A review of publications from leading pharmaceutical companies between 2014 and 2018 relating the use of AI in drug discovery revealed Random Forest (RF), SVMs, and other regression algorithms to be the most widely used ML techniques.[7]

Conversely, unsupervised learning algorithms are trained on input data that are not labeled and are often used as a part of exploratory data analysis, such as clustering and dimensionality reduction. Some standard unsupervised learning algorithms include kappa-means clustering (k-mean), hierarchical clustering, principal component analysis (PCA), self-organizing map (SOM), and t-distributed stochastic neighbor embedding (t-SNE). In general, these methods are termed dimensionality

reduction techniques and provide a means for projecting high-dimensional data into a low-dimensional space for visualization. They are often used in drug design for developing QSAR models (PCA-MLR modeling), designing screening libraries, clustering, data exploration, and visualizing the chemical space of large compound libraries. Several supervised learning algorithms, such as SVM and neural networks (NNs), can also support unsupervised learning.[33]

Unlike supervised and unsupervised learning, RL systems continuously interact with the environment using feedback from previous actions and experiences to achieve their goals. Each time a RL agent performs an action, it utilizes an objective function that is rewarded if the output is acceptable, and penalized when it is not. The goal of a RL algorithm is to identify the optimal policy to maximize the reward function. RL algorithms, such as Generative Tensorial Reinforcement Learning (GENTRL).[37] and Reinforcement Learning for Structural Evolution (ReLeaSE), have been used for designing molecules with desired properties during generative modeling.[38]

DL is a subset of ML and belongs to a broader family of artificial neural network (ANN) algorithms. It is currently the state-of-art AI technology and can be described as a class of representation-learning techniques. ANN algorithms are loosely inspired by the structure of the human brain. Accordingly, ANN architecture contains many processing elements, called neurons, that are organized into multiple layers. The network comprises input nodes and a layer of output nodes connected by layer of hidden nodes (Fig. 2d). Each hidden node has an associated weight, activation function, and bias function that transforms the input data to make predictions. The adjective 'deep' in DL refers to an ANN with many layers, and the number of hidden layers signifies the depth of the network. DL methods contains several hidden layers as opposed to traditional 'shallow-learning' ML methods, which usually contain one or two hidden layers. Whereas DL methods use deep and specialized architectures to learn and extract higher-level features in an automated fashion from unstructured data, they also require a huge amount of training data. Another key difference between DL and shallow-learning ML algorithms is that DL algorithms scale with the data, whereas shallow-learning ML algorithms converge at a certain level of performance. Some popular DL architectures used in drug discovery include convolutional neural networks (CNNs), graph CNNs (GCNNs), autoencoders (AE), and recurrent NNs (RNNs).

CNNs are the most-utilized DL method in drug discovery.[7] A CNN architecture comprises of multiple layers of neurons, each of which is fully connected to all neurons in the preceding layer. It usually contains several convolution layers and pooling layers occurring in an alternating fashion that are capable of learning any highly nonlinear function. Deep CNN models trained on 3D atomic grids extracted from experimental protein–ligand complex structures are now being used for structure-based VS and property prediction.[34–35] They have been shown to successfully model the complex, nonlinear phenomenon of a small molecule binding to the protein,[39] and have demonstrated significant improvement for property prediction.[40] Other DL architectures, such as RNNs, have also been used for generative modeling and focused library generation.[41]

Successful applications of AI in drug discovery

Over the past few years, several reviews have been published that highlight the emerging role of AI in drug discovery.[5,42–43] Hence, this review would focus on highlighting few exemplary studies where AI has made a real impact in small-molecule drug discovery.

Structural enablement of drug targets and binding site comparisons

The availability of atomic-resolution structural information of small molecules binding to drug targets provides opportunities for structure-guided hit identification (structure-based VS), fragment screening (fragment-based drug discovery; FBDD), and ligand optimization (structure-based drug design; SBDD). Structural information of the target also provides insights into selectivity drivers, resistance mechanisms, mode of action, allosteric pocket identification, and ligandability assessment for novel drug targets.[44–45] Despite technological advancements in X-ray crystallography, NMR spectroscopy, and single-particle cryo-EM, there is structural coverage for only ~ 35% of the human proteome. In many instances, this structural coverage is often limited to a single structural domain of a multidomain protein. Hence, there is still a gap between the number of known protein sequences and the number of experimentally solved structures.[46] Importantly, structural coverage for pharmaceutically relevant protein target families, such as G-protein-coupled receptors (GPCRs) and ion channels, remains underrepresented in databases, such as the Protein Data Bank (PDB).[47]

An alternative approach to protein 3D structure generation in the absence of experimental structure is the use of computational structure prediction methods. Homology modeling has been the traditional approach to bridge the sequence–structure gap. It predicts the 3D structure of an unknown (target) protein based on the experimental structure of a homologous (template) protein given its amino acid sequence. It has been shown that homology-modeled structures with sequence identity down to 30% are generally suitable for SBDD.[48] For proteins that lack a homologous structure, accurate structure prediction remains a challenge; however, advances in DL-based methods and the integration of coevolution data into modeling have invigorated the field of protein structure prediction.[49] DL-based algorithms, such as CNN, RNN, variational autoencoders (VAEs), and generative adversarial networks, have demonstrated improved success in protein structure prediction even in the absence of a template structure.[50]

The use of DL methods for protein structure prediction took center stage with the remarkable success of a deep convolutional residual network (ResNet)-based program called AlphaFold2 at the CASP14 competition.[51] AlphaFold2, developed by DeepMind technologies,[52] used a DNN architecture trained on 170 000 protein structures from the PDB to predict the distribution of distances between pairs of amino acids and torsion angles between chemical bonds that connect those amino acids in a protein. In addition, it uses evolutionary information derived from multiple sequence alignment and an end-to-end folding method for structure prediction. The methodology and the architecture behind AlphaFold2 was recently published.[53]

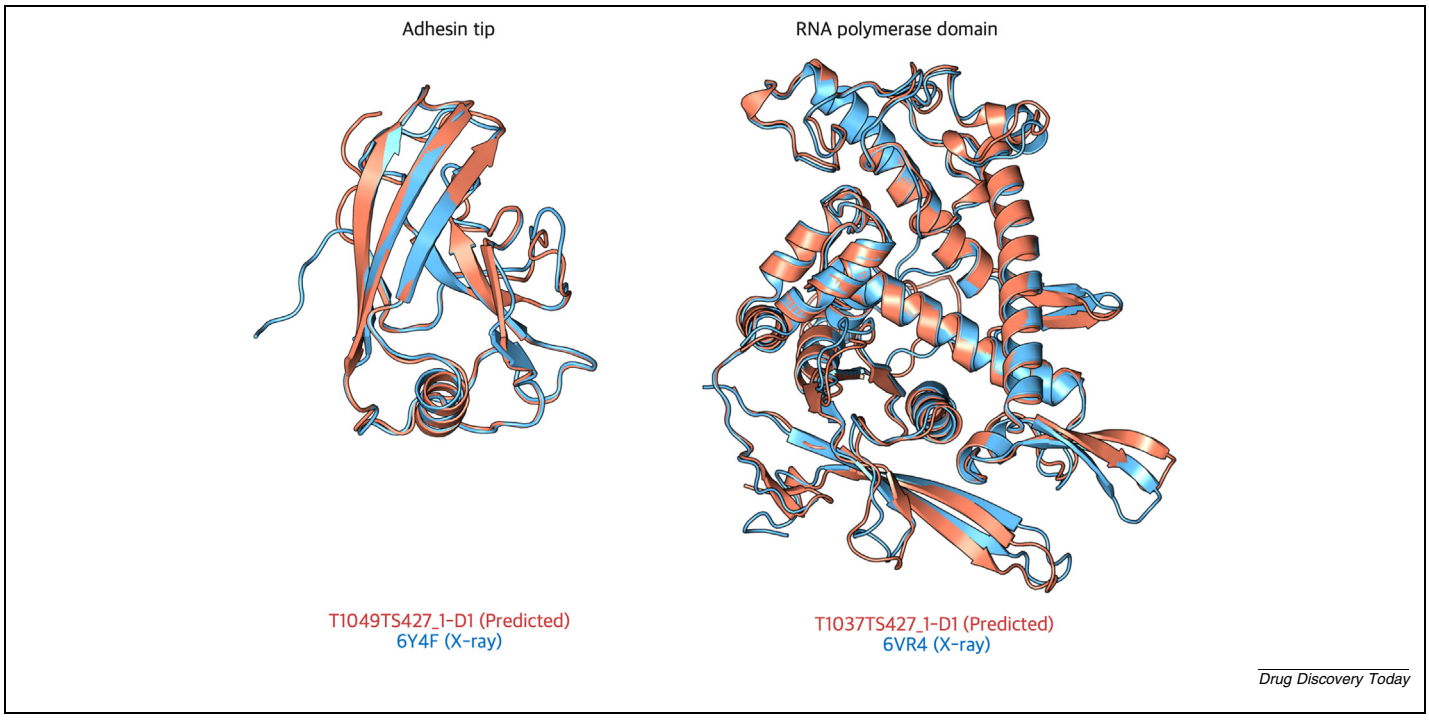


FIGURE 3
Structural superposition of experimental (red) and AlphaFold2 predicted (blue) protein structures.

To gauge structure prediction accuracy, CASP uses a Global Distance Test (GDT) metric that quantifies the residue correspondence between the model and the experimental structure. A score of 90 GDT implies the prediction accuracy to be comparable with experimental methods (Fig. 3); AlphaFold2 achieved a median score of 92.4 GDT for all targets. Thus, the results from the CASP14 competition revealed that DL methods could achieve impressive levels of accuracy comparable to experimental structures. DeepMind, in partnership with EMBL-EBI, has made freely available to the scientific community the 3D structures predicted by AlphaFold2, which offer structural coverage for 98.5% of the human proteome.[54] Inspired by the idea and the success of AlphaFold2, an academic team led by David Baker also developed a three-track NN program dubbed RoseTTAFold.[55]

Although these developments signify advancements in protein structure prediction, it is too early to state that AI has cracked the protein-folding problem or its impact on drug discovery will be transformative. AlphaFold2 was trained on over 170 000 protein structures from PDB, and any learning model is only as good as the data it was trained on. Current estimates of the number of folds in the PDB based on the SCOP version 2 database[56] is 1388, whereas the number of folds in nature is predicted to somewhere between 4000–10 000.[57] Hence, there are many novel folds, topologies, and architectures unseen in PDB, and there is also considerable redundancy in both sequence and protein families in this database. Furthermore, predicting the structure of multidomain proteins, multimeric protein complexes, and membrane proteins might be a harder problem to solve using AI. Nevertheless, DeepMind's technology promises to advance structural biology and *de novo* protein design, and to drive drug discovery.

Comparing protein-binding pockets at the structural genomics scale is a valuable exercise in structure-based drug design.[58] It provides information that can help understand selectivity, predict off-target liabilities, provide insights into drug repurposing, and aid protein function annotation. Traditional pocket comparison methods use representations, such as graph theory, geometric hashing, typed triangles, spherical harmonics, and physicochemical properties of the binding site atoms, to compute the sequence-free similarity between binding sites. These intuition-based featurization schemes can introduce human biases and are generally not scalable across thousands of binding sites.

The introduction of DNN algorithms has enabled the building of powerful voxel-based feature representations that can encode molecular properties and vectorize the binding sites into descriptor vectors. One such implementation is DeeplyTough,[59] which uses 3D steerable CNNs for comparing binding sites in an alignment-free manner. This is accomplished by encoding 3D representations of protein pockets into descriptor vectors, which can be used for computing pairwise Euclidean distances to quantifying pocket similarity. It was trained on the TOUGH-M1 data set, which is a nonredundant and representative data set of small-molecule binding pockets with approximately 1 million data points. It includes a positive subset comprising different proteins that bind chemically similar ligands, and a negative subset, containing different proteins that bind chemically dissimilar ligands. Performance was evaluated against two independently constructed datasets (Vertex and ProSPECCTs), with DeeplyTough demonstrating results competitive with existing approaches but with a reduced runtime.

Augmenting virtual screening using AI

VS is a computational technique that offers a complementary and cost-effective approach for hit identification relative to experimental screening methods, such as high-throughput screening.[60–63] Instead of physically screening every compound from the screening collection, VS uses computational techniques to prioritize a subset of compounds for evaluation in a primary assay.

The increasing size of 'make-on-demand' screening libraries, and the expanding number of high-value, challenging drug targets identified from functional genomics screening present a significant challenge for conventional VS techniques. Hence, AI methods that augment VS approaches and help efficiently explore the chemical space for hit identification have garnered considerable attention in drug discovery.

Ligand-based virtual screening

Ligand-based VS (LBVS) techniques aim to identify active compounds from a chemical library based on the principle of molecular similarity. They include similarity searching, pharmacophore screening, shape matching, and predictive modeling.

Predictive modeling for VS is an extension of the classical QSAR modeling paradigm. Classical QSAR uses statistical data-modeling methods on a congeneric series to build explanatory models that quantify SAR trends in a retrospective manner. Access to a large volume of chemogenomics data (PubChem's BioAssay[64] and the ChEMBL database[65]) and advances in ML and DL algorithms that can handle large data sets have provided new opportunities for QSAR modeling as a VS technique. Consequently, many successful applications of QSAR-based VS workflows for hit identification have been reported. Zhang *et al.* described the successful implementation of an ML-based QSAR workflow for VS that led to the discovery of novel antimalarial agents.[66] The authors used two ML algorithms (SVM and kNN) to develop a binary classifier model (active or inactive), which was trained using 3133 compounds with known antimalarial activities. The QSAR models were used to carry out VS against the ChemBridge database and resulted in the selection of 174 compounds for a follow-up screening in *Plasmodium falciparum* growth inhibition and cellular assays. Experimental validation revealed 25 of the selected compounds to be active, yielding a hit rate of 14.2%, with the most potent hit having an EC_{50} value of 95.6 nM. Subsequently, many studies have reported the application of ML and DL-based QSAR workflows as promising VS tools.[67–73]

Over the past decade, there has been a shift to web-based cheminformatics workbenches that streamline and automate ML- and DL-based QSAR workflows for VS. Liu *et al.* developed a user-friendly open-source web server called DeepScreening,[74] which allows users to build and validate RNN models using either ChEMBL bioactivity data or user-provided data sets for VS. DeepScreening also provides prebuilt DNN models for 1251 targets based on the bioactivity data curated from ChEMBL 24. This user-friendly interface and the availability of prebuilt QSAR models allow QSAR experts and nonexperts to perform VS against a specific target of interest. DpubChem[75] is another open-source web server that uses ML approaches to derive categorical QSAR models using PubChem data.

Whereas chemogenomic databases such as PubChem and ChEMBL provide sufficient bioactivity data to build models, there are still significant pitfalls associated with using these resources. Primary among these issues is the presence of bioactivity data from heterogeneous sources and an imbalanced ratio of active to inactive compounds for a given target. This makes the generalized use of QSAR-based workflows for VS more difficult to implement using public data sets compared with other virtual screening methods.

Structure-based virtual screening

A common computational strategy applied in SBDD is molecular docking. It plays a crucial role in driving many structurally enabled drug discovery programs, spanning from hit identification to LO and binding mode prediction.[76] The docking process involves predicting the bound ligand conformation (pose prediction) within a binding site, followed by an estimation of its binding affinity (scoring).[77]

Given significant progress in DL, the application of AI methods is becoming more commonplace in SBDD. Unlike shallow-learning AI methods that rely on feature engineering, DL can automatically learn and extract features from 3D structural data.[78] Thus, DL methods, which are popular in image recognition, are now being applied to extract and generalize structural features from protein–ligand complexes through multilayer feature extraction. This opens opportunities for using AI methods in SBVS, binding mode prediction, and binding affinity estimation. The use of ML and DL algorithms for pose prediction and scoring during docking has been demonstrated to have superior performance in terms of scoring power (ability to rank order compounds based on binding affinities), docking power (ability to distinguish native poses from decoy poses), and screening power (ability to distinguish binders from nonbinders).[79]

AtomNet® was the first structure-based application that used a deep CNN framework to predict binding affinity.[80] It uses a 3D grid approach to encode the environment of each atom in the binding site into voxelized feature vectors and was trained on a ChEMBL data set containing 78 000 actives and 2 000 000 decoys, spanning 290 targets. Interestingly, benchmark studies carried out using the DUD-E benchmark data set showed impressive performance, with AtomNet® achieving an area under the curve (AUC) greater than 0.9 on 57.8% of the targets, far surpassing the conventional docking methods.

Accurate binding mode prediction of the ligand using docking is a key challenge in SBDD. Although binding free energy is a macroscopic observable that involves a ratio of partition functions between two states (bound and unbound),[81] it is commonly assumed that the binding mode observed using experimental techniques, such as X-ray crystallography, corresponds to the lowest energy pose. Hence, most classical scoring functions that are parameterized to recapitulate binding affinities are in turn used to select the top-scoring docked pose as its predicted binding mode. However, accurate binding energy estimation using scoring functions remains a significant challenge, and this often results in 'hard failures', implying that the predicted binding modes do not correspond to native or near-native binding modes.[71]

The first attempt to develop a task-specific scoring function for binding pose prediction using ML was carried out by Ash-

tawy and Mahapatra.[82] They showed that various ML algorithms that can map structural and physicochemical information from protein–ligand complexes could distinguish native and near-native docked poses from decoy poses. The best task-specific ML scoring function displayed improved docking power (>14%) over classical scoring functions. A similar implementation that uses a 3D-CNN, called DeepBSP,[83] can predict the root mean square deviation (RMSD) of a predicted pose relative to its native binding pose. The authors trained their model against a data set that contained 11 925 native complexes and more than 165 000 Autodock Vina[84] docked decoy poses, and benchmarked their model using the CASF-2016 benchmark data set. Their findings revealed that scoring Autodock Vina generated poses with DeepBSP displayed an improved docking power relative to the hybrid knowledge and empirical-based scoring function available within the application.

In a prospective context, Adeshina *et al.* used 'vScreenML' for VS that led to the discovery of an AChE inhibitor with an IC_{50} value of 280 nM (K_i = 173 nM).[85] vScreenML is built on an XGBoost framework and uses a classifier approach to categorize docked poses as actives or decoys. The authors attributed its performance to the unique nature of the training data set, D-COID, which includes both native complexes that are representative of drug-like compounds and decoy complexes generated by molecular modeling. The inclusion of decoy complexes in the training data significantly improved the classification accuracy of the vScreenML scoring function.

Recently, a new class of scoring functions, developed based on ML and DL algorithms, has been gaining popularity.[17,86] Many common ML and DL architectures have been used for the development of ML-based scoring functions, including SVM, RF, kNN, gradient-boosting decision tree (GBDT), and 3D deep-CNN. Representative examples include Pafnucy,[87] Onion-Net,[88] RFScore-v3,[86] NNscore2.0,[89] BgN(BsN)-Score,[82] and Δ_{vina}RF.[90] Onion-Net and Pafnucy use a 3D CNN approach for featurization of protein–ligand complexes. ML-based scoring functions have been shown to outperform classical scoring functions in various benchmark studies.[17,91] Unlike traditional scoring functions, ML-based scoring functions are nonparametric because they do not have a predetermined functional form that approximates the underlying physics of molecular recognition.[17] Instead, they are trained from experimental data sets that contain both protein–ligand structural data and binding affinity data, such as PDBbind[92] and Binding MOAD.[93] Hence, they implicitly account for molecular interactions that are had to model explicitly. ML-based scoring functions can be used either to rescore docked poses generated by an external docking program or integrated within a docking program to help in guiding the sampling of poses. A detailed discussion of ML-based scoring functions is beyond the scope of this review, and the reader is referred to several excellent reviews on this topic.[17,94]

In addition, there are several studies that report the use of ML and DL-based scoring functions that lead to the identification of experimentally validated hits during VS campaigns.[17,89] As an example, an ML-based scoring function, called MIEC-SVM, which combines molecular interaction energy components (MIEC) and SVM, was used to screen the Specs vendor database and led to the identification of a novel class of ALK kinase inhibitor.[95] The use of MIEC-SVM to rescore Autodock 4.2 poses

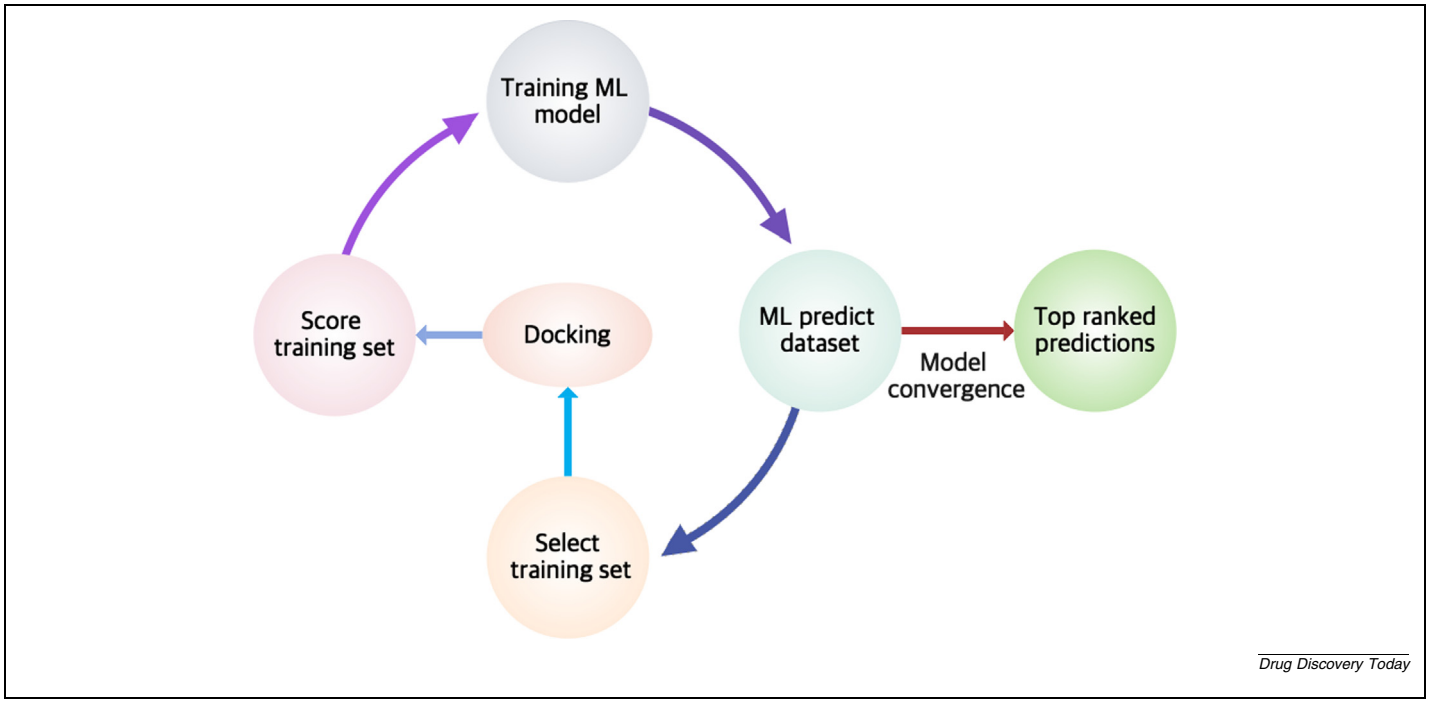


FIGURE 4

Overview of an active learning-based docking workflow. The active learning process begins with the selection of training data to train a machine learning (ML) model using the docking score. This procedure is iterated until the model converge and top-ranked molecules are recommended for further analysis (see main text for details). Figure adapted from.[19]

yielded a hit rate of 14% compared with a hit rate of 6% using the native scoring used by Autodock.

Although ML-based scoring functions offer an improvement over classical scoring functions, the interpretability of ML-based scoring functions is not straightforward[96] and one needs to be cautious of issues during benchmarking and validation of ML-based scoring functions. Flaws in the design of benchmark data sets and improper splitting of the data set into training and test data can yield overly optimistic performance estimates. Data leakage happens if the benchmark data set happens to have the information present in the training data leading to inflated performance during validation. Overfitting of ML occurs when models show high accuracy in the training set but fail to generalize on unseen data set. Gabel *et al*. reported on two ML-based scoring functions, RF-IChem and SVM-Ichem, which were used to predict pK_i values for 195 diverse protein–ligand X-ray structures.[97] The ML-based scoring functions were found to be insensitive to the position, orientation, and conformation of the docked pose, suggesting that they had overestimated accuracies from overfitting. Conversely, the Surflex-Dock scoring function behaved as expected, being logically sensitive to changes in pose conformation. This reminds us of the downside of developing computational tools with 'black box' characteristics and the potential to introduce artifacts due to overfitting.

Active learning docking

As the size of make-on-demand libraries continues to grow, so does the need for computational tools that can efficiently navigate the chemical space during VS. Although the debate whether 'bigger is better' or 'smaller is better' in VS continues,[98] docking of ultralarge libraries for hit identification is gaining popularity.[63,99] The colossal size of such libraries poses a challenge for docking programs, which can rarely perform brute force docking exceeding 100 million molecules.

Integrating an active learning algorithm with molecular docking offers an elegant solution for scaling up the screening of ultralarge libraries.[18] Active learning docking, in general, begins by docking a small subset of the entire library, then uses the results to train an ML model to predict docking scores for the rest of the compounds in the library. Top-scoring compounds from the ML model are then docked, and the ML model is updated with the new data. The whole process is repeated iteratively until the ML model converges. Thus, active learning docking, while preserving the fidelity of brute force docking, helps identify the highest-scoring compounds from ultralarge compound libraries at a fraction of the time required by conventional docking. An overview of this process is shown in Fig. 4.

One such implementation of active learning docking is 'Active Learning Glide',[19] which integrates the docking program Glide[100] and the ML algorithms available from the open-source framework DeepChem.[101] Graff *et al*. demonstrated that a molecular-pool based active learning-guided docking approach was able to retrieve most of the top-scoring compounds in a virtual library at a fraction of the computational cost of brute force docking.[18]

Generative chemistry

The concept of using computational methods for compound ideation has a rich history.[102] Early structure-based *de novo* design approaches involved automated and incremental construction of ligands within the receptor binding site. Programs such as LUDI identify potential interaction sites in the binding pocket and build molecules from a predefined set of organic fragments that sterically and electronically complement the protein binding pocket.[103] Inverse-QSAR modeling is another *de novo* molecular design approach that seeks to design molecules with a desired activity or property by inversely mapping the molecular descriptor from a preconstructed quantitative structure–activity/property relationship (QSAR/QSPR) models. In general, solving the inverse-QSAR problem is complex, because reconstructing molecular structures based on molecular descriptor information provided by a forward QSAR/QSPR model is challenging. A practical problem that precluded the widespread adoption of these classical *de novo* design approaches is the limited synthesizability and poor drug-like properties of the designed molecules.

Over the past few years, the use of AI-based generative modeling algorithms for *de novo* molecular design has gained in popularity[104–105] because they can overcome issues encountered with classical *de novo* design approaches. Generative chemistry relies on the use of modern AI-based generative modeling tools to generate synthetically tractable compounds with drug-like properties while satisfying the desired target property profile. Based on a data-driven approach, generative modeling algorithms learn the underlying nonlinear distribution between molecular structures, their biological activity, and physicochemical properties from a large volume of data to inform compound design. AI-guided generative modeling platforms, in short, perform compound ideation, prediction, and selection of compounds with favorable properties. Several DL architectures, such as VAEs, generative adversarial networks (GANs), RL, and RNNs, have been applied to *de novo* molecular design.[105–107] Current generative modeling methodologies can also be categorized depending on the underlying method used for molecular featurization. Whereas most first-generation generative modeling methods used fingerprints and SMILES strings to encode molecular structures, newer approaches, such as molecular graphs and fragment-based methods, are becoming increasingly popular.

One such study to establish proof-of-principle for implementing deep GANs in generative modeling was reported by Kadurin *et al*.[108] An example demonstrating the utility of RNNs in generative modeling was also reported by Segler *et al*..[109] The first report of the successful application of RNN containing long short-term memory (LSTM) cells for *de novo* molecular design was reported by Gupta *et al*..[41] The generative LSTM model was trained against the ChEMBL22 database to generate novel molecules that could modulate retinoid X receptors (RXRs)[110] and peroxisome proliferator-activated receptors (PPARs).[111] The fine-tuning process involved training against a data set containing 25 RXR and PPAR modulators (agonists and partial agonists). The on-target activity of the generated compounds was predicted using an ML model, with four of the five top-ranked compounds selected for synthesis showing activity in a cell-based assay (Fig. 5a, 1–4). Two were found to be PPAR agonists, and two were dual inhibitors of PPAR and RXR, displaying EC_{50} values ranging from low to double-digit mM. Although the compounds were not extensively characterized, these findings demonstrate the

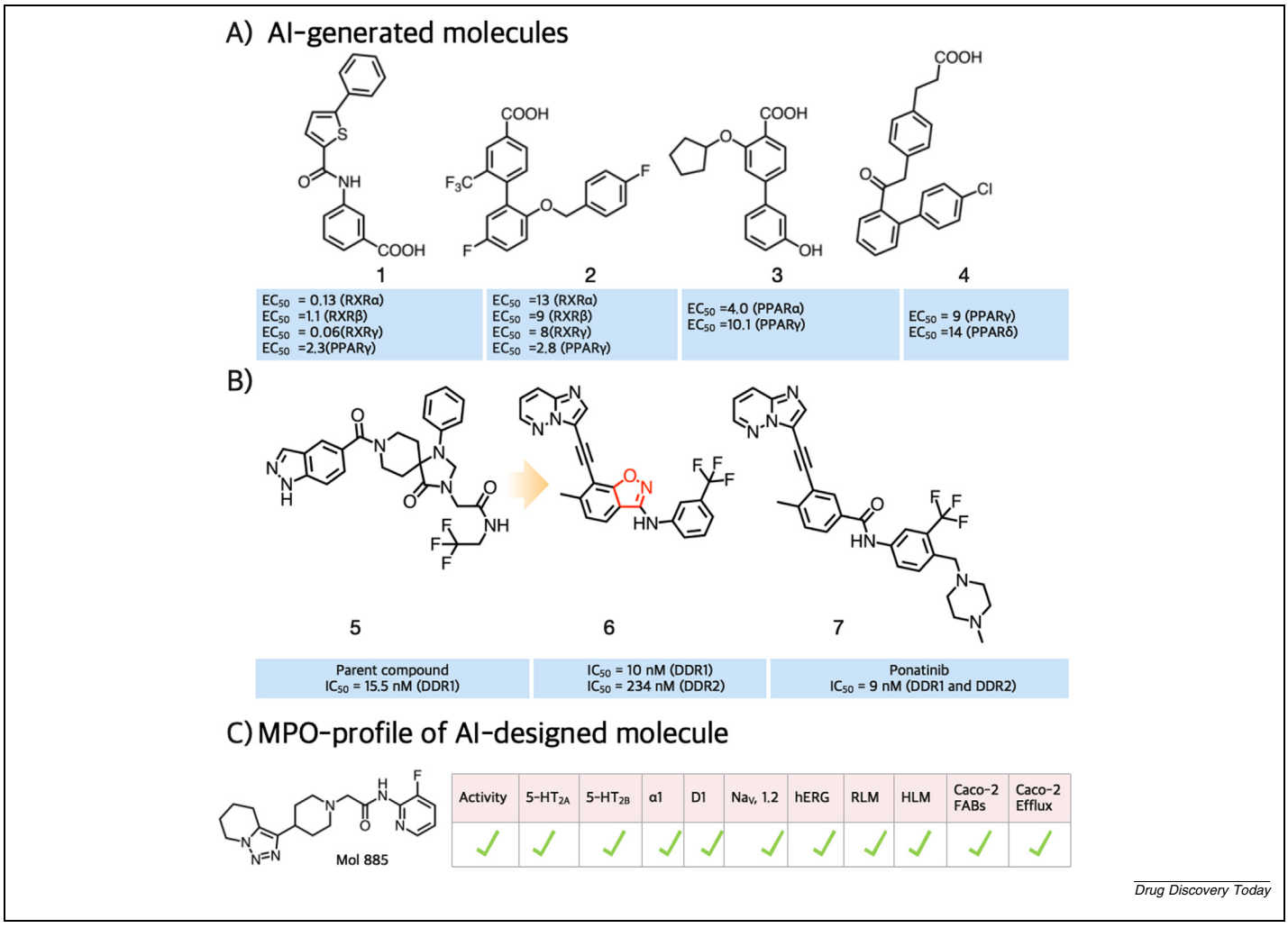


FIGURE 5

Chemical structures of the molecules generated by artificial intelligence (AI)-based generative modeling. **(a)** *De novo* design of retinoid X receptor (RXR) and peroxisome proliferator-activated receptor (PPAR) inhibitors using long short-term memory (LSTM) deep learning (DL) (**1**–**4**).[110] **(b)** The parent compound (**5**) a discoidin domain receptor 1 (DDR1) inhibitor (US 10,239,876 B2) served as a starting point for *de novo* design. The chemical structure of an AI-generated molecule (**6**) from Zhavoronkov *et al.*[37] compared with ponatinib (**7**), a marketed multikinase inhibitor. **(c)** The multiparameter optimization (MPO) profile of AI-based generative modeling designed molecule (Mol 885).[114]

ability of generative AI to deliver synthetically tractable, novel bioactive molecules that satisfy design objectives

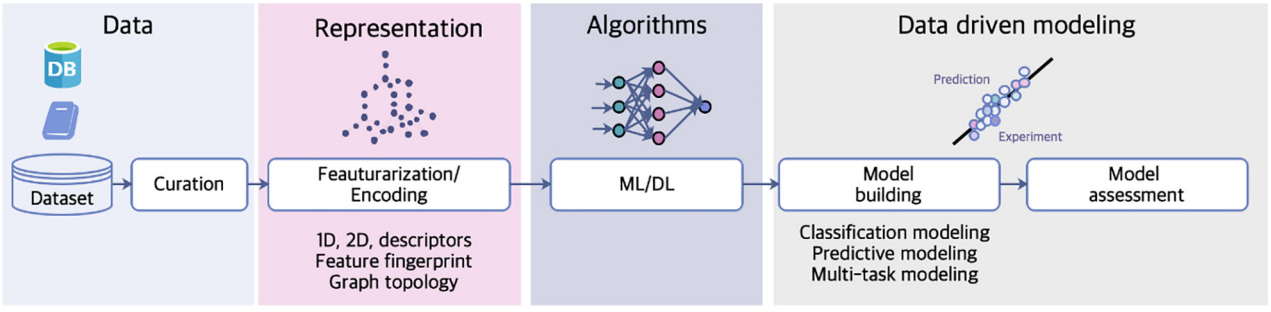
Another generative modeling study that attracted significant press coverage at the time was the use of a GENTRL model by Zhavoronkov *et al.*, which led to the discovery of potent kinase discoidin domain receptor (DDR1) inhibitors in only 21 days.[37] The molecule (**6**) designed by the GENTRL approach is shown in Fig. 5b and is compared with the parent molecule (**5**) and other DDR1 inhibitor (**7**). The authors trained their generative model in a semisupervised fashion using an objective function that rewards synthetic feasibility, on-target activity, and novelty. Of the 30 000 molecules proposed by the generative model, six were subsequently synthesized and tested. Four compounds were found to be active in biochemical assays, and two were active in a cell-based assay, with the best compound (**6**) having an IC_{50} value close to 10 nM in both biochemical and cell-based assays. Although this study demonstrated the ability of generative modeling to identify a nanomolar hit compound, concerns have been raised about the novelty of the molecules.[112] The best molecule designed was strikingly similar to the marketed multityrosine kinase inhibitor ponatinib. In addition, selectivity profiling of the compound against the broader kinome was not established, which called into question the clinical value of the compound. The authors responded to these criticisms, stating that the study was meant to demonstrate the potential of generative modeling technology and not intended at identifying a clinical candidate compound.[113]

Recently, Perron *et al.* gave an account of what might be the first report of a successful application of generative modeling in solving a MPO problem.[114] An LSTM generative model coupled to a RL method, on an undisclosed target, was used to design 150 compounds that were predicted to meet all the defined LO criteria. The training data set included 881 molecules with 11 sets of associated assay data, including on-target activity, off-target activity, and ADMET endpoints. None of the compounds in the training data satisfied all the property and potency criteria.

A)

Absorption	Distribution	Metabolism	Excretion	Toxicity	PhysChem
Caco-2 permeability	Human serum albumin binding	Microsomal stability	CL_{int}	Ames mutagenicity	Aq. solubility
Caco-2 efflux ratio	Plasma-protein binding	Hepatocyte stability	Terminal half-life	CYP inhibition	LogD
PAMPA permeability	Fraction unbound (fu)	Site of metabolism		Phaspholipidosis	Membrane affinity
Blood-brain barrier penetration	Volume of distribution (vdss)			hERG inhibition	pKa
	Brain/plasma ratio (Kp)			Drug-induced liver injury (DILI)	

B)



Drug Discovery Today

FIGURE 6

In silico absorption, distribution, metabolism, excretion, and toxicity (ADMET) prediction. **(a)** ADMET and PhysChem endpoints for which *in silico* models are available. **(b)** Overview of a data-driven model building workflow.

Twenty compounds generated by the model were shortlisted for synthesis based on different criteria; nine of these compounds failed during synthesis, the remaining 11 were synthesized and profiled, and one of these satisfied all the eleven LO criteria (Fig. 5c).

Although generative chemistry is becoming increasingly popular, emphasis should also be placed on the rigorous validation of generative models. Assessment methods for generative modeling should include the application of distribution learning benchmarks, synthetic validity, novelty, compound quality, goal-directed objectives, as a part of their evaluation framework. Open-source standardized benchmarking platforms, such as Molecular Sets (MOSES)[115] and GuccaMol,[116] could serve as a framework for benchmarking generative modeling methods.

In silico ADMET prediction

The observation that poor pharmacokinetics of drug candidates were an important cause of clinical attrition in the late 1990s brought about a paradigm shift within the pharmaceutical industry.[117] It witnessed the emergence of several property-based drug-likeness rules, such as Lipinski's Ro5,[118] and many developability metrics to control compound properties during LO.[119–121] In addition, the establishment of miniaturized, high-throughput *in vitro* ADMET profiling assays resulted in the parallel evaluation of efficacy and ADMET during earlier stages of drug discovery. *In silico* ADMET modeling is intended to assist project teams in the design and selection of novel compounds with superior ADMET properties and in directing experimental resources to the most favorable compounds, thereby reducing the overall number of compounds that need to be synthesized and profiled.[122–123] Over the years, pharmaceutical companies have deployed many global *in silico* ADMET models in their discovery pipelines. A representative list of such *in silico* ADMET models is provided in Fig. 6a.

Early work in ADMET modeling used linear regression methods, such as those used by Hansch[124] and Free-Wilson analysis.[125] However, with the development of ML algorithms and the availability of large-scale homogenous ADMET data,[126] *in silico* ADMET modeling transitioned toward ML-based predictive models developed using Bayesian neural networks, RFs, and SVMs. These ML algorithms are suitable for predicting endpoints that have a complex and nonlinear relationship.

The use of DNN methods for *in silico* modeling of ADMET endpoints gained popularity following the Kaggle 'Merck Molecular Activity Challenge' conducted in 2012.[40] The Kaggle competition was meant to examine how well ML methods can predict 18 different ADMET endpoints using data sets of various sizes (2000–50 000 molecules) derived from Merck's internal data. The winning entry used an ensemble approach that included DNN, gradient-boosting machine (GBM), and Gaussian process (GP) regression methods.[127] Merck researchers released a follow-up study comparing the performance of DNN models to that of RF models, and demonstrated that DNN models outperformed RF in most cases. Similarly, in the Tox21 data challenge, conducted by the NIH to compare computational models for toxicity prediction, DL models excelled and outperformed shallow-learning ML models.[128]

A unique characteristic of DNNs is their ability to simultaneously train NNs that combine different endpoints within a single model. Using a learning technique called inductive transfer

learning, multitask DNN trains data sets corresponding to different ADMET endpoints and combines them into a single model. The rationale behind multitask DNNs is to enable faster learning and improved model accuracy by sharing their representation internally.[129] Most multitask DNN models used for modeling ADMET endpoints use a 'Hard' parameter sharing approach, which implies sharing of the hidden layer between all the tasks.[130] A general data-driven ADMET model building process is illustrated in Fig. 6b.

In an accuracy performance benchmark study carried out against 31 assay data sets, Evan *et al*. showed that multitask DNNs were more accurate in predicting ADMET endpoints over single-task DNNs and shallow-learning ML methods, such as RF.[131] Scientists from Sanofi–Aventis AG also reported the successful implementation of predictive multitask DNN model into their *in silico* ADMET workflows.[132] By applying an alternate multitask learning method to transfer features between data sets, they developed multitask models for predicting metabolic liability/clearance, Caco-2 permeability, and $\log D_{7.4}$. They also reported that multitask DNN methods, although outperforming single-task DNN methods in many instances, showed poorer performance compared with single-task DNNs for certain endpoints. Hence, combining mechanistically unrelated endpoints in a multitask model could lead to poor performance, since the information shared between the tasks might not be correlated.[133] Hence, *a priori* assumption of the predictive advantage of multitask DNN over single-task DNNs is a challenge and both approaches need to be evaluated when developing predictive ADMET models.

Although predictive modeling has a significant role in selecting compounds with superior ADMET properties, they are often insufficient for guiding compound ideation during the lead optimization phase. Extracting tacit knowledge from the corpus of information generated from prior discovery programs can be used as a source of idea for compound design. One such widely used concept in medicinal chemistry is termed Matched Molecular Pair (MMP) analysis.[134]

MMP is defined as a pair of molecules that differ only by a well-defined structural transformation that is associated with a relative change in a property value.[135] Traditionally, MMP analysis was carried out by analyzing the frequency of one-to-one structural transformations. With advances in AI technology and molecule fragmentation algorithms, MMP analysis of large data sets in an automated fashion is now feasible. A notable implementation is MCPairs,[136] which uses an unsupervised ML approach to mine *in vitro* ADMET data integrated from three different pharmaceutical companies (AstraZeneca, Genentech, and Roche). The use of AI and the availability of large-scale data help develop next-generation MMP platforms that offer practical solutions to address ADMET issues using explainable AI.

Computer-aided synthesis planning

The use of computer-aided synthetic planning (CASP) dates back to the pioneering work of E.J. Corey,[137] who formalized the concept of 'retrosynthetic analysis' during the late 1960s.[138] Retrosynthetic analysis refers to a technique that involves the deconstruction of a target molecule into its simple, readily available starting materials by sequential disconnections and functional group interconversions. CASP programs incorporate the idea of retrosynthetic analysis and help synthetic organic chemists select the most efficient and cost-effective synthetic route.[24,139] They can also be used for selectivity and side product prediction and for recommending and evaluating reaction conditions. The use of AI has revitalized the field of computer-aided synthesis planning, and technological developments over the years have been well reviewed in recent literature.[24,140] Therefore, we limit the scope of our review to the application of CASP in the context of drug discovery. AI-assisted synthesis planning tools help chemists augment their synthetic chemistry knowledge by recommending viable synthetic routes. They also help chemists make better decisions, thereby improving efficiency and productivity by reducing synthesis failures.[141] Ultimately, this accelerates the 'make' phase of the DMTA cycle in drug discovery.

Computer-aided synthetic route-planning strategies generally fall into two broad categories: rule or template-based methods and template-free methods. Rule-based methods use expert coded rules and heuristics extracted from reaction databases and literature for suggesting synthetic routes. In rule-based methods, the reaction rules are extracted and codified manually. One such example of retrosynthetic software, which uses a library of expert encoded rules for chemical synthesis planning, is Synthia (formerly Chematica). A limitation with such rule-based methods is its inability to scale with the exponential growth of chemical literature, and its knowledge base is limited because complete coverage is unlikely. To address these limitations, automated rule-based methods were developed for extracting reaction rules from reaction data sets using computational approaches. Such automated, rule-based methods use template extraction algorithms that rely on atom-mapped reaction examples in the form of SMIRKS patterns for extracting transformations from reaction data sets. Two significant limitations include the high computational cost involved with the calculation of subgraph isomorphism and the lack of chemical intelligence.[142–143] An alternate rule-based approach developed in recent years for extracting reaction rules is the application of data-driven DL techniques. A notable example is a pioneering study by Segler *et al*.[109] who used a neural-symbolic approach to extract retrosynthetic rules from the Reaxys database autonomously without expert input. These rules were then subjected to reaction prediction in combination with a modern Monte-Carlo tree search algorithm to select the most promising retrosynthetic steps.

An orthogonal approach to the classical rule-based methods is the use of template-free methods for reaction prediction and retrosynthetic transformations. Template-free methods draw inspiration from Natural Language Processing (NLP) and treat forward or retrosynthetic prediction as a neural machine translation problem.[144] Given that molecules can be represented as SMILES strings, each chemical reaction can be encoded as sentences and treated as a chemical linguistics translational problem. Liu and coworkers proposed the first template-free model for retrosynthetic analysis.[145] They used a sequence-to-sequence (Seq2-Seq) model that used encoder–decoder-based natural NLP transformers to map a SMILES representation of the reactants to a SMILES representation of the respective products and vice versa. The NN architecture employed uses bidirectional LSTM

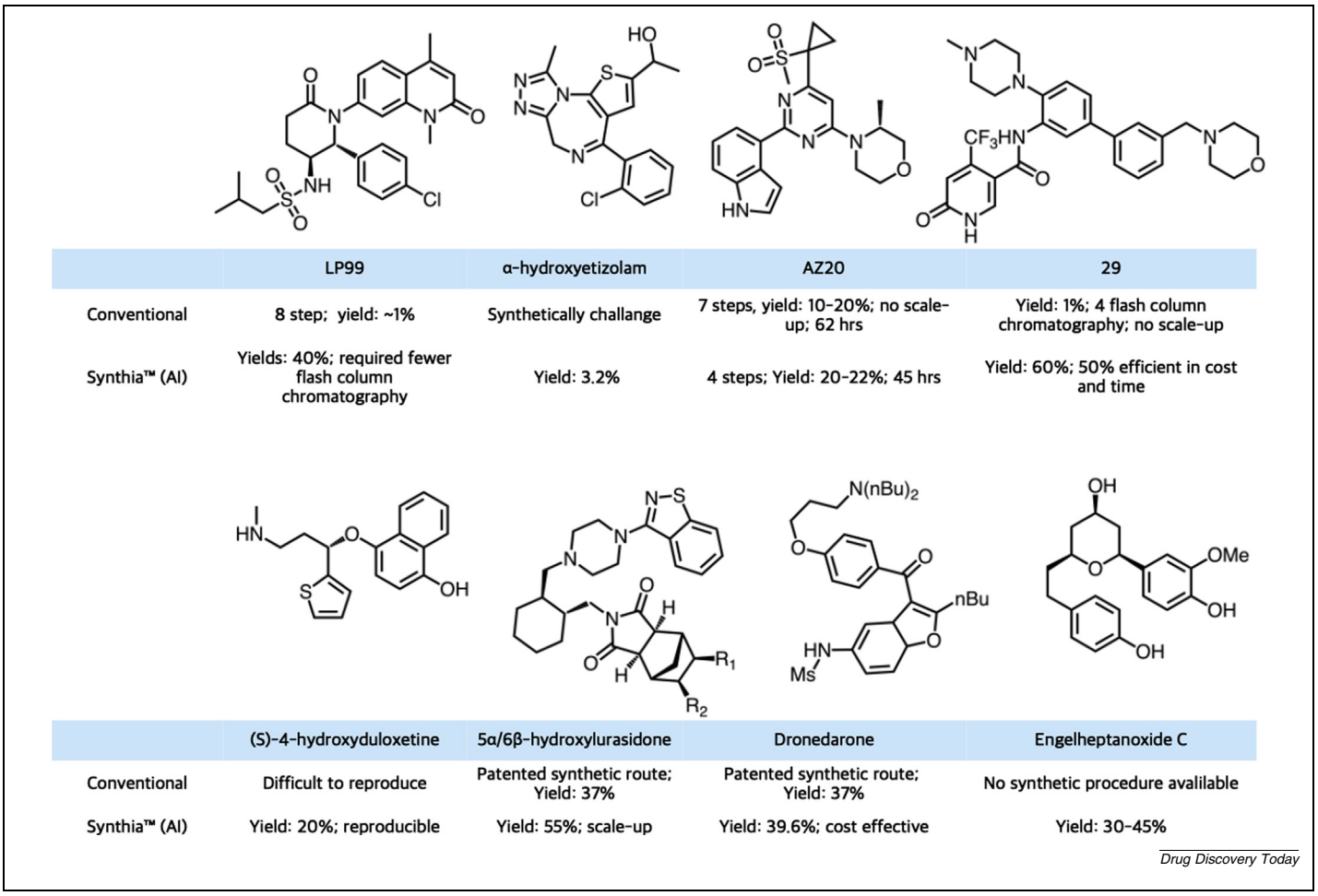


FIGURE 7
Synthetic efficiency comparison for eight target molecules synthesized based on conventional and computer-aided synthesis planning (CASP) routes proposed by Synthia™ (formerly known as Chematica).

cells with an additive attention mechanism for token-wise alignment. This model was shown to be comparable to rule-based expert systems in solving retrosynthetic reaction prediction tasks. Other template-free approaches that have also been reported to show promising results include the use of a graph, chemical reaction networks, and similarity-based methods.[146–147] Some popular retrosynthetic planning tools include AiZynthFinder,[148] Spaya.ai (https://spaya.ai), and the Chemistry42™ platform.[149]

The first report describing the successful execution of a multistep synthesis route proposed by a synthesis planning software was disclosed in 2018 by Klucznik *et al.*[150] The authors used the Synthia™ software to design synthetic pathways for eight structurally diverse and synthetically challenging target molecules (Fig. 7). Synthia™, which relies on a library of ~ 50 000 expert encoded reaction rules compiled over a period of 15 years, was able to propose synthetic routes within 15–20 min for all targets. Interestingly, the synthetic routes proposed by Synthia™ were significantly different from their original synthetic routes disclosed in the patents and provided higher yields in fewer synthetic steps.

Additionally, synthesis planning tools have opened the possibility of enumerating and exploring the synthetically accessible chemical space. An early example is the creation of the reaction-driven Pfizer Global Virtual Library (PGVL).[151] Enumeration of a synthetically accessible chemical library involves the use of reaction-based enumeration tools that use well-established reactions and exclusion rules, proprietary chemical 'know-how' protocols, and building block availability information. The Enamine REAL Space library (https://enamine.net) is one prime example of a 'make-on-demand' library, comprising ~ 15.5 billion compounds at the last count, making it the largest commercially available library of its kind. Although AI has shown great promise in streamlining synthetic organic chemistry, there are still opportunities for further improvements, such as the reliable prediction of stereochemical outcomes, reagent prediction, reaction conditions, and so forth.

Challenges in the application of AI to preclinical drug discovery

AI has generated a wave of excitement and investment within the biopharmaceutical industry. Although proponents of AI technology believe that it will usher in a new era of AI-driven drug discovery, skeptics argue that most of the promises are tantalizing and aspirational. However, most experts agree the reality

will likely fall somewhere between the two.[5] Although there have been several notable advances demonstrating the impact of AI in preclinical drug discovery projects, it is not yet clear how far we are from an era of AI-driven drug discovery. Currently, AI in preclinical drug discovery is riding the 'peak of inflated expectations' phase of the of Gartner's hype cycle.[107] Hence, it is essential to sift hyperbole from reality and set realistic expectations.

There are many challenges in implementing AI into drug discovery, one of the most demanding of which is the requirement of large amounts of high-quality training data. Building a useful and predictive model for decision-making largely depends on both the quality of the data and data set size. Unfortunately, data generation in drug discovery is both resource intensive and time consuming, often leading to a compound profiling strategy that measures few endpoints during the early stages that covers many project compounds, followed by intensive profiling of a small number of compounds during the late stages to support progression. This is evident from the millions of bioactivity data points available in commercial and public databases, such as GOSTAR, PubChem, and ChEMBL, but relatively fewer associated ADMET data points, leaving us with an incomplete data matrix. Recent advances, such as sparse AI methods and deep imputation methods, could help mitigate data paucity issues.

Standardization and integration of the available drug discovery data also present a challenge during data curation. Assay readouts are often expressed in different formats (e.g., IC_{50}, EC_{50}, K_d, K_i, or % inhibition) that are not readily compatible, and the underlying data types can be discrete or continuous. Furthermore, the readouts are assay specific and comparable only under certain conditions, because they could differ in the assay format, protein construct length, substrate concentration, and so on. Hence, integrating and standardizing both public and proprietary data to expand the usable data volume is a significant challenge.

Uncertainty in drug discovery data is pervasive, and an estimate of the experimental uncertainty of K_i values for compounds with multiple activity values reported in ChEMBL had revealed a mean unsigned error (MUE) of 0.44 pKi units.[113] The inherent noise associated with experimental data is referred to as aleatoric uncertainty[12] and the uncertainty value in the training data sets the upper limit of performance that a predictor model can achieve. In addition, drug discovery data often span a small dynamic range (2–3 log units), which often limits model predictivity.

The chemical space encompassed by an ML/DL model is termed the 'applicability domain' (AD) and this space is necessarily minuscule compared with the available chemical space, estimated to be $\sim 10^{60}$ molecules.[152] In this way, every conceivable global model is a local model, and predictions carried for compounds outside the AD of the model are generalizations based on inductive inference, which increases the uncertainty in the predictions for such compounds.

Another limitation of using ML and DL algorithms is the lack of transparency, because they operate as 'black boxes', implying that the features, functions, and weights encoded by these models are beyond the interpretation of a human user. The opaque decision-making process used by these algorithms cannot help discovery scientists make prospective suggestions. Thus, such opaque models are used to predict the properties of compounds already designed, limiting the utility of the models to support decision-making. Thus, for compound ideation, explainable AI (XAI), which provides transparent, informative, and interpretable findings to drive compound ideation, is needed. However, modern XAI algorithms are now emerging that assist with the interpretation of these black-box models. These algorithms use approaches such as sensitivity analysis, variable importance, and partial derivatives to extract those variables or substructural features used by the model for prediction.

Preclinical data rely on using proxy measures, such as cellular target engagement as a predictor of human *in vivo* target occupancy, patient-derived xenograft (PDX) mouse tumor models as a proxy for clinical efficacy in oncology programs, human HepG2 cells as a surrogate for genotoxicity, and the Caco-2 cell permeability assay as a surrogate for estimating human intestinal permeability. Although preclinical data help support clinical translation, these surrogate or proxy *in vivo* data points cannot be reliably used to train AI models to predict clinical outcomes (e.g., human PK, clinical efficacy, safety, and tolerability).[153]

Apart from the scientific and technical challenges stated above, organizational culture and agility are also crucial for the adoption and implementation of AI. There is certainly some level of entrenchment and a greater degree of risk aversion in implementing new technologies in many organizations. Implementing AI technologies, to stay ahead, while realizing there is no guaranteed success, would require making some bold decisions and vision from senior leadership.[154]

Concluding remarks

The application of AI technologies holds great promise for brining down drug discovery timelines and cost. Although AI might not be a panacea for every problem in drug discovery, it is clearly a valuable tool if applied in the correct context and with the right data. The strength of AI technologies will certainly be used to complement human intelligence and augment our capabilities, perhaps changing the way we approach drug discovery, but not as a replacement for human ingeneuity. Drawing parallels between Google's DeepMind besting a human professional Go player with drug discovery is incongruous; drug discovery is high-dimensional science displaying nonlinearity with many known unknowns and unknown unknowns, not a Go game that can be defined based on a finite set of rules. Although one needs to be receptive in embracing new technologies such as AI for drug discovery, healthy skepticism and caution are advisable as the field matures.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that might be perceived as affecting the perception of this review.

Acknowledgments

The authors are grateful to Drs. Sreetama Das, Jun Pei, and Vigneshwaran Namasivayam for their valuable comments and discussions. **Disclaimer:** Views expressed in this article are

solely those of the authors and do not necessarily reflect the view of affiliated institutions.

References

1 O.J. Wouters, M. McKee, J. Luyten, Research and development costs of new drugs–reply, JAMA 324 (2020) 518.
2 A. Mullard, 2020 FDA drug approvals, Nat Rev Drug Discov 20 (2021) 85–90.
3 D.G. Brown, H.J. Wobst, A decade of FDA-approved drugs (2010–2019): trends and future directions, J Med Chem 64 (2021) 2312–2338.
4 N. Fleming, How artificial intelligence is changing drug discovery, Nature 557 (2018) S55–S57.
5 P. Schneider, W.P. Walters, A.T. Plowright, N. Sieroka, J. Listgarten, R.A. Goodnow Jr, et al., Rethinking drug design in the artificial intelligence era, Nat Rev Drug Discov 19 (2020) 353–364.
6 Properzi, F, Steedman, M, Taylor, K, Ronte, H, Haughey, J. (2019) Intelligent drug discovery powered by AI. Deloitte Insights; www2.deloitte.com/us/en/insights/industry/life-sciences/artificial-intelligence-biopharma-intelligent-drug-discovery.html [Accessed November 19, 2021]..
7 A. Schuhmacher, A. Gatto, M. Kuss, O. Gassmann, M. Hinder, Big techs and startups in pharmaceutical R&D - a 2020 perspective on artificial intelligence, Drug Discov Today 26 (2021) 2226–2231.
8 T. Aoyama, Y. Suzuki, H. Ichikawa, Neural networks applied to pharmaceutical problems. III. Neural networks applied to quantitative structure-activity relationship (QSAR) analysis, J Med Chem 33 (1990) 2583–2590.
9 D.E. Klingler, Expert systems in the pharmaceutical industry, Drug Info J 22 (1988) 249–258.
10 R.M. Plenge, Disciplined approach to drug discovery and early development, Sci Transl Med 8 (2016) 349PS315.
11 S. De Cesco, J.B. Davis, P.E. Brennan, TargetDB: a target information aggregation tool and tractability predictor, PLoS ONE 15 (2020) e0232644.
12 J. Jiménez-Luna, F. Grisoni, G. Schneider, Drug discovery with explainable artificial intelligence, Nat Mach Intell 2 (2020) 573–584.
13 P.A. Vignaux, E. Minerali, D.H. Foil, A.C. Puhl, S. Ekins, Machine learning for discovery of gsk3beta inhibitors, ACS Omega 5 (2020) 26551–26561.
14 K. McCloskey, E.A. Sigel, S. Kearnes, L. Xue, X. Tian, D. Moccia, et al., Machine learning on DNA-encoded libraries: a new paradigm for hit finding, J Med Chem 63 (2020) 8857–8866.
15 D.B. Kitchen, H.Y. Decornez, Computational techniques to support hit triage, in: W. Czechtizky, P. Hamley (Eds.), Small Molecule Medicinal Chemistry, John Wiley & Sons, Inc., Hoboken, 2015, pp. 189–220.
16 C. Stork, Y. Chen, M. Sicho, J. Kirchmair, Hit dexter 2.0: Machine-learning models for the prediction of frequent hitters, J Chem Inf Model 59 (2019) 1030–1043.
17 Q.U. Ain, A. Aleksandrova, F.D. Roessler, P.J. Ballester, Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening, Wiley Interdiscip Rev Comput Mol Sci 5 (2015) 405–424.
18 D.E. Graff, E.I. Shakhnovich, C.W. Coley, Accelerating high-throughput virtual screening through molecular pool-based active learning, Chem Sci 12 (2021) 7866–7881.
19 Y. Yang, K. Yao, M.P. Repasky, K. Leswing, R. Abel, B.K. Shoichet, S.V. Jerome, Efficient exploration of chemical space with docking and deep-learning, J Chem Theory Comput 17 (2021) 7106–7119.
20 S.M. Paul, D.S. Mytelka, C.T. Dunwiddie, C.C. Persinger, B.H. Munos, S.R. Lindborg, et al., How to improve R&D productivity: the pharmaceutical industry's grand challenge, Nat Rev Drug Discov 9 (2010) 203–214.
21 M.D. Segall, Multi-parameter optimization: identifying high quality compounds with a balance of properties, Curr Pharm Des 18 (2012) 1292–1310.
22 M.M. Hann, G.M. Keseru, Finding the sweet spot: the role of nature and nurture in medicinal chemistry, Nat Rev Drug Discov 11 (2012) 355–365.
23 Y. Bian, X.Q. Xie, Generative chemistry: drug discovery with deep learning generative models, J Mol Model 27 (2021) 71.
24 C.W. Coley, W.H. Green, K.F. Jensen, Machine learning in computer-aided synthesis planning, Acc Chem Res 51 (2018) 1281–1289.
25 D. Webb, T.F. Jamison, Continuous flow multi-step organic synthesis, Chem Sci 1 (2010) 675–680.
26 B.W.J. Irwin, T.M. Whitehead, S. Rowland, S.Y. Mahmoud, G.J. Conduit, M.D. Segall, Deep imputation on large-scale drug discovery data, Applied AI Letters 2 (2021) e31.
27 T.M. Whitehead, B.W.J. Irwin, P. Hunt, M.D. Segall, G.J. Conduit, Imputation of assay bioactivity data using deep learning, J Chem Info Model 59 (2019) 1197–1204.
28 B.W.J. Irwin, J.R. Levell, T.M. Whitehead, M.D. Segall, G.J. Conduit, Practical applications of deep learning to impute heterogeneous drug discovery data, J Chem Inf Model 60 (2020) 2848–2857.
29 H. Dolgos, M. Trusheim, D. Gross, J.P. Halle, J. Ogden, B. Osterwalder, et al., Translational medicine guide transforms drug development processes: the recent Merck experience, Drug Discov Today 21 (2016) 517–526.
30 T.S. Toh, F. Dondelinger, D. Wang, Looking beyond the hype: applied AI and machine learning in translational medicine, EBioMedicine 47 (2019) 607–615.
31 S. Ekins, A.C. Puhl, K.M. Zorn, T.R. Lane, D.P. Russo, J.J. Klein, et al., Exploiting machine learning for end-to-end drug discovery and development, Nat Mater 18 (2019) 435–441.
32 J. Vamathevan, D. Clark, P. Czodrowski, I. Dunham, E. Ferran, G. Lee, et al., Applications of machine learning in drug discovery and development, Nat Rev Drug Discov 18 (2019) 463–477.
33 Y.C. Lo, S.E. Rensi, W. Torng, R.B. Altman, Machine learning in chemoinformatics and drug discovery, Drug Discov Today 23 (2018) 1538–1546.
34 L.O. Jimenez, D.A. Landgrebe, Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data, IEEE Trans Syst Man Cybern, Part C (Applications and Reviews) 28 (1998) 39–54.
35 A. Golbraikh, X.S. Wang, H. Zhu, A. Tropsha, Predictive qsar modeling: Methods and applications in drug discovery and chemical risk assessment, in: J. Leszczynski (Ed.), Handbook of Computational Chemistry, Springer, Netherlands, Amsterdam, 2012, pp. 1309–1342.
36 I.H. Witten, E. Frank, M.A. Hall, C.J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann:, Burlington, 2016.
37 A. Zhavoronkov, Y.A. Ivanenkov, A. Aliper, M.S. Veselov, V.A. Aladinskiy, A.V. Aladinskaya, et al., Deep learning enables rapid identification of potent ddr1 kinase inhibitors, Nat Biotechnol 37 (2019) 1038–1040.
38 M. Popova, O. Isayev, A. Tropsha, Deep reinforcement learning for de novo drug design, Sci Adv 4 (2018) eaap7885.
39 Gomes, J, Ramsundar, B, Feinberg, E.N, Pande, V.S. Atomic convolutional networks for predicting protein–ligand binding affinity. *arXiv 2017:* arXiv:1703.10603 2017..
40 Kaggle Merck molecular activity challenge, Kaggle Inc. www.Kaggle.Com/c/merckactivity [Accessed November 19, 2021]..
41 A. Gupta, A.T. Muller, B.J.H. Huisman, J.A. Fuchs, P. Schneider, G. Schneider, Generative recurrent networks for de novo drug design, Mol Info 37 (2018) 170011.
42 M.A. Sellwood, M. Ahmed, M.H. Segler, N. Brown, Artificial intelligence in drug discovery, Future Med Chem 10 (2018) 2025–2028.
43 D. Paul, G. Sanap, S. Shenoy, D. Kalyane, K. Kalia, R.K. Tekade, Artificial intelligence in drug discovery and development, Drug Discov Today 26 (2021) 80–93.
44 F.N. Edfeldt, R.H. Folmer, A.L. Breeze, Fragment screening to predict druggability (ligandability) and lead discovery success, Drug Discov Today 16 (2011) 284–287.
45 S. Vukovic, D.J. Huggins, Quantitative metrics for drug-target ligandability, Drug Discov Today 23 (2018) 1258–1266.
46 C. UniProt, Uniprot: the universal protein knowledgebase in 2021, Nucleic Acids Res 49 (2021) D480–D489.
47 H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, et al., The Protein Data Bank, Nucleic Acids Res 28 (2000) 235–242.
48 A. Hillisch, L.F. Pineda, R. Hilgenfeld, Utility of homology models in the drug discovery process, Drug Discov Today 9 (2004) 659–669.
49 B. Kuhlman, P. Bradley, Advances in protein structure prediction and design, Nat Rev Mol Cell Biol 20 (2019) 681–697.
50 L.N. Kinch, A. Kryshtafovych, B. Monastyrskyy, N.V. Grishin, CASP 13 target classification into tertiary structure prediction categories, Proteins 87 (2019) 1021–1036.
51 Jumper, J, Evans, R, Pritzel, A, Green, T, Figurnov, M, Tunyasuvunakool, K. *et al.* (2020) High accuracy protein structure prediction using deep learning. In *Fourteenth Critical Assessment of Techniques for Protein Structure Prediction*. Davis: Protein Structure Prediction Center; 2020: 22–24..
52 The AlphaFold Team, Alphafold: a solution to a 50-year-old grand challenge in biology. https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in–biology [accessed on November 19, 2021]..

53 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, et al., Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021) 583–589.
54 K. Tunyasuvunakool, J. Adler, Z. Wu, T. Green, M. Zielinski, A. Zidek, et al., Highly accurate protein structure prediction for the human proteome, Nature 596 (2021) 590–596.
55 M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G.R. Lee, et al., Accurate prediction of protein structures and interactions using a three-track neural network, Science 373 (2021) 871–876.
56 A. Andreeva, E. Kulesha, J. Gough, A.G. Murzin, The scop database in 2020: expanded classification of representative family and superfamily domains of known protein structures, Nucleic Acids Res 48 (2020) D376–D382.
57 A. Grant, D. Lee, C. Orengo, Progress towards mapping the universe of protein folds, Genome Biol 5 (2004) 107.
58 C. Ehrt, T. Brinkjost, O. Koch, Impact of binding site comparisons on medicinal chemistry and rational molecular design, J Med Chem 59 (2016) 4121–4151.
59 M. Simonovsky, J. Meyers, DeeplyTough: learning structural comparison of protein binding sites, J Chem Inf Model 60 (2020) 2356–2366.
60 T. Zhu, S. Cao, P.C. Su, R. Patel, D. Shah, H.B. Chokshi, et al., Hit identification and optimization in virtual screening: practical recommendations based on a critical literature analysis, J Med Chem 56 (2013) 6560–6572.
61 P. Ripphausen, B. Nisius, L. Peltason, J. Bajorath, Quo vadis, virtual screening? A comprehensive survey of prospective applications, J Med Chem 53 (2010) 8461–8467.
62 K.L. Damm-Ganamet, N. Arora, S. Becart, J.P. Edwards, A.D. Lebsack, H.M. McAllister, et al., Accelerating lead identification by high throughput virtual screening: prospective case studies from the pharmaceutical industry, J Chem Inf Model 59 (2019) 2046–2062.
63 C. Gorgulla, A. Boeszoermenyi, Z.F. Wang, P.D. Fischer, P.W. Coote, K.M. Padmanabha Das, et al., An open-source drug discovery platform enables ultra-large virtual screens, Nature 580 (2020) 663–668.
64 Y. Wang, S.H. Bryant, T. Cheng, J. Wang, A. Gindulyte, B.A. Shoemaker, et al., PubChem bioassay: 2017 update, Nucleic Acids Res 45 (2017) D955–D963.
65 D. Mendez, A. Gaulton, A.P. Bento, J. Chambers, M. De Veij, E. Felix, et al., ChEMBL: towards direct deposition of bioassay data, Nucleic Acids Res 47 (2019) D930–D940.
66 L. Zhang, D. Fourches, A. Sedykh, H. Zhu, A. Golbraikh, S. Ekins, et al., Discovery of novel antimalarial compounds enabled by QSAR-based virtual screening, J Chem Inf Model 53 (2013) 475–492.
67 M. Anantpadma, T. Lane, K.M. Zorn, M.A. Lingerfelt, A.M. Clark, J.S. Freundlich, et al., Ebola virus Bayesian machine learning models enable new *in vitro* leads, ACS Omega 4 (2019) 2353–2361.
68 X. Chen, W. Xie, Y. Yang, Y. Hua, G. Xing, L. Liang, et al., Discovery of dual FGFR4 and EGFR inhibitors by machine learning and biological evaluation, J Chem Inf Model 60 (2020) 4640–4652.
69 M.J. Donlin, T.R. Lane, O. Riabova, A. Lepioshkin, E. Xu, J. Lin, et al., Discovery of 5-nitro-6-thiocyanatopyrimidines as inhibitors of *Cryptococcus neoformans* and *Cryptococcus gattii*, ACS Med Chem Lett 12 (2021) 774–781.
70 Z. Liu, D. Huang, S. Zheng, Y. Song, B. Liu, J. Sun, et al., Deep learning enables discovery of highly potent anti–osteoporosis natural products, Eur J Med Chem 210 (2021) 112982.
71 G.M. Verkhivker, D. Bouzida, D.K. Gehlhaar, P.A. Rejto, S. Arthurs, A.B. Colson, et al., Deciphering common failures in molecular docking of ligand-protein complexes, J Comput Aided Mol Des 14 (2000) 731–751.
72 K.J. Wicht, J.M. Combrinck, P.J. Smith, T.J. Egan, Bayesian models trained with HTS data for predicting beta-haematin inhibition and *in vitro* antimalarial activity, Bioorg Med Chem 23 (2015) 5210–5217.
73 M. Yang, B. Tao, C. Chen, W. Jia, S. Sun, T. Zhang, et al., Machine learning models based on molecular fingerprints and an extreme gradient boosting method lead to the discovery of JAK2 inhibitors, J Chem Inf Model 59 (2019) 5002–5012.
74 Z. Liu, J. Du, J. Fang, Y. Yin, G. Xu, L. Xie, Deepscreening: a deep learning-based screening web server for accelerating drug discovery, Database (Oxford) 2019 (2019) baz104.
75 O. Soufan, W. Ba-Alawi, A. Magana-Mora, M. Essack, V.B. Bajic, DPubChem: a web tool for qsar modeling and high-throughput virtual screening, Sci Rep 8 (2018) 9110.
76 D.B. Kitchen, H. Decornez, J.R. Furr, J. Bajorath, Docking and scoring in virtual screening for drug discovery: methods and applications, Nat Rev Drug Discov 3 (2004) 935–949.
77 I. Halperin, B. Ma, H. Wolfson, R. Nussinov, Principles of docking: an overview of search algorithms and a guide to scoring functions, Proteins 47 (2002) 409–443.
78 J.C. Pereira, E.R. Caffarena, C.N. Dos Santos, Boosting docking-based virtual screening with deep learning, J Chem Inf Model 56 (2016) 2495–2506.
79 Z. Wang, H. Sun, X. Yao, D. Li, L. Xu, Y. Li, et al., Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power, Phys Chem Chem Phys 18 (2016) 12964–12975.
80 Wallach, I, Dzamba, M, Heifets, A. AtomNet: a deep, convolutional neural network for bioactivity prediction in structure-based drug discovery. *arXiv* 2015: arXiv:1510.02855v1.
81 M.K. Gilson, J.A. Given, B.L. Bush, J.A. McCammon, The statistical-thermodynamic basis for computation of binding affinities: a critical review, Biophys J 72 (1997) 1047–1069.
82 H.M. Ashtawy, N.R. Mahapatra, Machine-learning scoring functions for identifying native poses of ligands docked to known and novel proteins, BMC Bioinformatics 16 (Suppl. 6) (2015) S3.
83 J. Bao, X. He, J.Z.H. Zhang, DeepBSP-a machine learning method for accurate prediction of protein-ligand docking structures, J Chem Inf Model 61 (2021) 2231–2240.
84 O. Trott, A.J. Olson, Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J Comput Chem 31 (2010) 455–461.
85 Y.O. Adeshina, E.J. Deeds, J. Karanicolas, Machine learning classification can reduce false positives in structure-based virtual screening, Proc Natl Acad Sci U S A 117 (2020) 18477–18488.
86 P.J. Ballester, J.B. Mitchell, A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking, Bioinformatics 26 (2010) 1169–1175.
87 M.M. Stepniewska-Dziubinska, P. Zielenkiewicz, P. Siedlecki, Development and evaluation of a deep learning model for protein-ligand binding affinity prediction, Bioinformatics 34 (2018) 3666–3674.
88 L. Zheng, J. Fan, Y. Mu, Onionnet: A multiple-layer intermolecular-contact-based convolutional neural network for protein-ligand binding affinity prediction, ACS Omega 4 (2019) 15956–15965.
89 J.D. Durrant, J.A. McCammon, NNScore 2.0: A neural-network receptor-ligand scoring function, J Chem Inf Model 51 (2011) 2897–2903.
90 C. Wang, Y. Zhang, Improving scoring-docking-screening powers of protein–ligand scoring functions using random forest, J Comput Chem 38 (2017) 169–177.
91 C. Shen, Y. Hu, Z. Wang, X.J. Zhang, H.Y. Zhong, G.A. Wang, et al., Can machine learning consistently improve the scoring power of classical scoring functions? Insights into the role of machine learning in scoring functions, Brief Bioinform 22 (2021) 497–514.
92 R. Wang, X. Fang, Y. Lu, C.Y. Yang, S. Wang, The PDBbind database: methodologies and updates, J Med Chem 48 (2005) 4111–4119.
93 M.L. Benson, R.D. Smith, N.A. Khazanov, B. Dimcheff, J. Beaver, P. Dresslar, et al., Binding MOAD, a high-quality protein-ligand database, Nucleic Acids Res 36 (2008) D674–D678.
94 H.J. Li, K.H. Sze, G. Lu, P.J. Ballester, Machine-learning scoring functions for structure-based virtual screening, Wiley Interdiscip Rev Comput Mol Sci 11 (2021) e1478.
95 H. Sun, P. Pan, S. Tian, L. Xu, X. Kong, Y. Li, et al., Constructing and validating high-performance miec–svm models in virtual screening for kinases: a better way for actives discovery, Sci Rep 6 (2016) 24817.
96 J. Sieg, F. Flachsenberg, M. Rarey, In need of bias control: evaluating chemical data for machine learning in structure-based virtual screening, J Chem Inf Model 59 (2019) 947–961.
97 J. Gabel, J. Desaphy, D. Rognan, Beware of machine learning-based scoring functions-on the danger of developing black boxes, J Chem Inf Model 54 (2014) 2807–2815.
98 D.E. Clark, Virtual screening: Is bigger always better? Or can small be beautiful?, J Chem Info Model 60 (2020) 4120–4123.
99 J. Lyu, S. Wang, T.E. Balius, I. Singh, A. Levit, Y.S. Moroz, et al., Ultra-large library docking for discovering new chemotypes, Nature 566 (2019) 224–229.
100 R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, et al., Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, J Med Chem 47 (2004) 1739–1749.
101 B. Ramsundar, P. Eastman, P. Walters, V. Pande, K. Leswing, Z. Wu, Deep Learning for the Life Sciences, O'Reilly Media, Sebastopol, 2019.
102 F. Darvas, Application of the sequential simplex method in designing drug analogs, J Med Chem 17 (1974) 799–804.
103 H.J. Bohm, Ludi: Rule-based automatic design of new substituents for enzyme inhibitor leads, J Comput Aided Mol Des 6 (1992) 593–606.

104 M. Elbadawi, S. Gaisford, A.W. Basit, Advanced machine-learning techniques in drug discovery, Drug Discov Today 26 (2021) 769–777.
105 J. Meyers, B. Fabian, N. Brown, De novo molecular design and generative models, Drug Discov Today 26 (2021) 2707–2715.
106 O. Prykhodko, S.V. Johansson, P.-C. Kotsias, J. Arús-Pous, E.J. Bjerrum, O. Engkvist, et al., A de novo molecular generation method using latent vector based generative adversarial network, J Cheminform 11 (2019) 74.
107 Q. Vanhaelen, Y.C. Lin, A. Zhavoronkov, The advent of generative chemistry, ACS Med Chem Lett 11 (2020) 1496–1505.
108 A. Kadurin, S. Nikolenko, K. Khrabrov, A. Aliper, A. Zhavoronkov, Drugan: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico, Mol Pharm 14 (2017) 3098–3104.
109 M.H.S. Segler, T. Kogej, C. Tyrchan, M.P. Waller, Generating focused molecule libraries for drug discovery with recurrent neural networks, ACS Cent Sci 4 (2018) 120–131.
110 D. Merk, F. Grisoni, L. Friedrich, G. Schneider, Tuning artificial intelligence on the de novo design of natural-product-inspired retinoid x receptor modulators, Commun Chem 1 (2018) 68.
111 D. Merk, L. Friedrich, F. Grisoni, G. Schneider, De novo design of bioactive small molecules by artificial intelligence, Mol Inform 37 (2018) 1700153.
112 W.P. Walters, M. Murcko, Assessing the impact of generative ai on medicinal chemistry, Nat Biotechnol 38 (2020) 143–145.
113 A. Zhavoronkov, A. Aspuru-Guzik, Reply to 'assessing the impact of generative ai on medicinal chemistry', Nat Biotechnol 38 (2020) 146.
114 Quentin, P, Olivier, M, Hamza, T, Adam, S, Anne, R, Arnaud, G. *et al*. Deep generative models for ligand-based de novo design applied to multi-parametric optimization. *ChemRxiv*. Published online January 25, 2021. http://dx.doi.org/10.26434/chemrxiv.13622417.v2..
115 D. Polykovskiy, A. Zhebrak, B. Sanchez-Lengeling, S. Golovanov, O. Tatanov, S. Belyaev, et al., Molecular sets (moses): a benchmarking platform for molecular generation models, Front Pharmacol 11 (2020) 565644.
116 N. Brown, M. Fiscato, M.H.S. Segler, A.C. Vaucher, Guacamol: benchmarking models for de novo molecular design, J Chem Inf Model 59 (2019) 1096–1108.
117 I. Kola, J. Landis, Can the pharmaceutical industry reduce attrition rates?, Nat Rev Drug Discov 3 (2004) 711–716.
118 C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv Drug Deliv Rev 46 (2001) 3–26.
119 R.J. Young, D.V.S. Green, C.N. Luscombe, A.P. Hill, Getting physical in drug discovery II: the impact of chromatographic hydrophobicity measurements and aromaticity, Drug Discov Today 16 (2011) 822–830.
120 Y.C. Martin, A bioavailability score, J Med Chem 48 (2005) 3164–3170.
121 T.T. Wager, X.J. Hou, P.R. Verhoest, A. Villalobos, Central nervous system multiparameter optimization desirability: Application in drug discovery, ACS Chem Neurosci 7 (2016) 767–775.
122 H. van de Waterbeemd, E. Gifford, ADMET in silico modelling: towards prediction paradise?, Nat Rev Drug Discov 2 (2003) 192–204.
123 Lombardo, F, Desai, P.V, Arimoto, R, Desino, K.E, Fischer, H, Keefer, C.E. *et al*. In silico absorption, distribution, metabolism, excretion, and pharmacokinetics (ADME-PK): utility and best practices. An industry perspective from the international consortium for innovation through quality in pharmaceutical development. *J Med Chem* 2017; 60: 9097–9113..
124 C. Hansch, P.P. Maloney, T. Fujita, R.M. Muir, Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficients, Nature 194 (1962) 178–180.
125 S.M. Free Jr, J.W. Wilson, A mathematical contribution to structure-activity studies, J Med Chem 7 (1964) 395–399.
126 A.H. Goller, L. Kuhnke, F. Montanari, A. Bonin, S. Schneckener, A. Ter Laak, et al., Bayer's in silico ADMET platform: a journey of machine learning over the past two decades, Drug Discov Today 25 (2020) 1702–1709.
127 J. Ma, R.P. Sheridan, A. Liaw, G.E. Dahl, V. Svetnik, Deep neural nets as a method for quantitative structure–activity relationships, J Chem Info Model 55 (2015) 263–274.
128 G.B. Goh, N.O. Hodas, A. Vishnu, Deep learning for computational chemistry, J Comput Chem 38 (2017) 1291–1307.
129 B. Ramsundar, B. Liu, Z. Wu, A. Verras, M. Tudor, R.P. Sheridan, et al., Is multitask deep learning practical for pharma?, J Chem Info Model 57 (2017) 2068–2076.
130 S. Sosnin, M. Vashurina, M. Withnall, P. Karpov, M. Fedorov, I.V. Tetko, A survey of multi-task learning methods in chemoinformatics., Mol Inform 38 (2019) e1800108.
131 E.N. Feinberg, E. Joshi, V.S. Pande, A.C. Cheng, Improvement in ADMET prediction with multitask deep featurization, J Med Chem 63 (2020) 8835–8848.
132 J. Wenzel, H. Matter, F. Schmidt, Predictive multitask deep neural network models for ADME-tox properties: learning from large data sets, J Chem Inf Model 59 (2019) 1253–1268.
133 Y. Xu, J. Ma, A. Liaw, R.P. Sheridan, V. Svetnik, Demystifying multitask deep neural networks for quantitative structure-activity relationships, J Chem Inf Model 57 (2017) 2490–2504.
134 Kenny, P.W, Sadowski, J. Structure modification in chemical databases. In: Oprea TI, ed. *Chemoinformatics in Drug Discovery*. Weinheim: Wiley–VHC Verlag GmbH & Co. KGaA; 2005: 271–285..
135 E. Griffen, A.G. Leach, G.R. Robb, D.J. Warner, Matched molecular pairs as a medicinal chemistry tool, J Med Chem 54 (2011) 7739–7750.
136 C. Kramer, A. Ting, H. Zheng, J. Hert, T. Schindler, M. Stahl, et al., Learning medicinal chemistry absorption, distribution, metabolism, excretion, and toxicity (ADMET) rules from cross-company matched molecular pairs analysis (MMPA), J Med Chem 61 (2018) 3277–3292.
137 E.J. Corey, A.K. Long, S.D. Rubenstein, Computer-assisted analysis in organic synthesis, Science 228 (1985) 408–418.
138 E.J. Corey, General methods for the construction of complex molecules, Pure Appl Chem 14 (1967) 19–38.
139 P.P. Plehiers, C.W. Coley, H. Gao, F.H. Vermeire, M.R. Dobbelaere, C.V. Stevens, et al., Artificial intelligence for computer-aided synthesis in flow: analysis and selection of reaction components, Front Chem Eng 2 (2020) 5.
140 Z. Wang, W. Zhao, G.F. Hao, B.A. Song, Mapping the resources and approaches facilitating computer-aided synthesis planning, Org Chem Front 8 (2021) 812–824.
141 A. Thakkar, S. Johansson, K. Jorner, D. Buttar, J.-L. Reymond, O. Engkvist, Artificial intelligence and automation in computer aided synthesis planning, React Chem Eng 6 (2021) 27–51.
142 S. Szymkuc, E.P. Gajewska, T. Klucznik, K. Molga, P. Dittwald, M. Startek, et al., Computer-assisted synthetic planning: the end of the beginning, Angew Chem Int Ed Engl 55 (2016) 5904–5937.
143 J. Dong, M. Zhao, Y. Liu, Y. Su, X. Zeng, Deep learning in retrosynthesis planning: datasets, models and tools, Brief Bioinform (2021) bbab391.
144 T.J. Struble, J.C. Alvarez, S.P. Brown, M. Chytil, J. Cisar, R.L. DesJarlais, et al., Current and future roles of artificial intelligence in medicinal chemistry synthesis, J Med Chem 63 (2020) 8667–8682.
145 B. Liu, B. Ramsundar, P. Kawthekar, J. Shi, J. Gomes, Q. Luu Nguyen, et al., Retrosynthetic reaction prediction using neural sequence-to-sequence models, ACS Cent Sci 3 (2017) 1103–1113.
146 C.W. Coley, L. Rogers, W.H. Green, K.F. Jensen, Computer-assisted retrosynthesis based on molecular similarity, ACS Cent Sci. 3 (2017) 1237–1245.
147 M. Sacha, M. Blaz, P. Byrski, P. Dabrowski-Tumanski, M. Chrominski, R. Loska, et al., Molecule edit graph attention network: modeling chemical reactions as sequences of graph edits, J Chem Inf Model 61 (2021) 3273–3284.
148 S. Genheden, A. Thakkar, V. Chadimová, J.-L. Reymond, O. Engkvist, E. Bjerrum, AiZynthFinder: a fast, robust and flexible open–source software for retrosynthetic planning, J Cheminform 12 (2020) 70.
149 Yan A. Ivanenkov, Y.A, Zhebrak, A, Bezrukov, D, Zagribelnyy, B, Aladinskiy, V, Polykovskiy, D, Putin, E, Kamya, P, Aliper, A, Zhavoronkov, A, Chemistry42: an AI-based platform for de novo molecular design. *arXiv* 2021: arXiv:2101.09050v1..
150 T. Klucznik, B. Mikulak-Klucznik, M.P. McCormack, H. Lima, S. Szymkuć, M. Bhowmick, et al., Efficient syntheses of diverse, medicinally relevant targets planned by computer and executed in the laboratory, Chem 4 (2018) 522–532.
151 Q. Hu, Z. Peng, S.C. Sutton, J. Na, J. Kostrowicki, B. Yang, et al., Pfizer global virtual library (PGVL): a chemistry design tool powered by experimentally validated parallel synthesis information, ACS Comb Sci 14 (2012) 579–589.
152 A. Mullard, The drug-maker's guide to the galaxy, Nature 549 (2017) 445–447.
153 A. Bender, I. Cortés-Ciriano, Artificial intelligence in drug discovery: What is realistic, what are illusions? Part 1: ways to make an impact, and why we are not there yet, Drug Discov Today 26 (2021) 511–524.
154 Mitra, S. The state of AI adoption in pharma. ZA, 2018. www.Zs.Com/insights/the-state-of-ai-adoption-in-pharma [Accessed November 19, 2021]..